

Private vs public genomics?

Commercial sequencing of genomes has stimulated scientific progress. But the increasing value of such companies risks exacerbating the conflict between the interests of investors and of the public. Both need to worry.

What was the purpose of the announcement by Celera Genomics on Monday that it now has DNA sequences totalling more than 90 per cent of the genome in its database? If the company's goal was to raise the value of its shares, it certainly succeeded. On the day of the announcement, the shares, which had ended the previous day of trading at US\$187, leapt at one stage by more than \$70, and ended the day at \$242 — giving the company, at least on paper, a market value of \$6.3 billion.

But that explanation is not totally persuasive. Celera's shares, like those of a raft of other genomics-based companies, have been rising steadily for several weeks, partly because of various scientific announcements, such as the successful sequencing of human chromosome 22, and also because some investors are turning to biotechnology following a spree with Internet stocks.

Yet neither is a purely scientific explanation convincing. The 90 per cent 'milestone' announced with pride by Celera's president (and prime mover) Craig Venter, and the fact that "greater than 97 per cent of all human genes" are now present in its database, are somewhat arbitrary benchmarks. Both are considerable achievements, and some of these data are no doubt already proving of significant value to the private subscribers to the database. But, as Venter himself admits, much remains to be done before anyone understands what the newly discovered genes represent and before the data can be assembled into anything resembling a complete genome sequence.

Whatever the motivation for the announcement, one of its most significant aspects was that it underlined the basic complementarity between the private and public sequencing efforts. Celera officials openly admit that the company has only been able to achieve the 90 per cent figure by drawing extensively on sequence data already made openly and immediately available through the Human Genome

Project (see page 119). As their statement puts it, "the combination of these two complementary genome sequencing and assembly approaches greatly reduces the time for Celera to finish the sequence of the human genome".

All the more disturbing, therefore, that Celera clings to the right to control the use of its own data, while freely using that obtained from publicly financed sources. Venter's comments that the full consensus sequence data, once published in a scientific journal, will be made "freely available" to researchers around the world, are welcome. But the company's caveat that this will only be done "under a non-redistribution agreement" raises the spectre that scientists seeking to use these data in, for example, the development of potential commercial products will find themselves constrained by strict conditions.

Celera has a right to benefit directly from its own heavy investment in sequencing, and such sequencing efforts could be as valuable to public sequencing initiatives as the latter have already been to Celera's. But Celera should take note of the proposed National Institutes of Health guidelines on the use of research materials (http://www.nih.gov/od/ott/RTguide_final.htm). Although not bound by the guidelines, Celera and other companies should heed the warning they give against excessive burdens of chains of licensing requirements.

Are the interests of investors and the public at odds? Celera's playing of the market increasingly threatens such a conflict: the financial expectations now being placed upon the genomics industry will necessitate secrecy and potentially burdensome licensing, whereas public funding bodies rightly seek the greatest possible access to, and freedom to build on, raw sequence data. Given the benefits of rapid development of treatments of disease, investors could find their interests set not only against those of academics and science funding agencies but also the public at large. Remember Monsanto? ■

Millennium bugs

AIDS in Africa threatens the United States. It should act accordingly.

The United Nations (UN) security council this week had, for the first time in its history, a health issue on its agenda: AIDS in Africa. Al Gore, who chaired the meeting, announced that the US administration had asked Congress for \$150 million for vaccine research and prevention programmes in Africa. Will the money ever materialize? Will it be new money? The dismal history of such pledges gives grounds for scepticism.

Nevertheless, and despite the horrific scale of the problems in Africa, there is room for hope. Three years ago, malaria was an invisible disease. Since then research organizations, donor agencies and drug companies have created a consortium to stimulate research and strategies to combat the disease, while companies have teamed up to relaunch drug discovery. These new structures and their plans have their flaws (see *Nature Medicine* 5,1334; 1999), but they testify to movement and commitment where before there was only stagnation. Similar groups of people will meet in Cape Town next month for a major meeting whose goal is to do the same for tuberculosis.

Media attention and political commitment are essential to provide the climate for such initiatives to flourish. Rightly, Al Gore has pointed to the absolute necessity of investment in basic health infrastructure and domestic competence, and to the need for African governments themselves to lift health from the bottom of their spending priorities. Wisely, he has characterized AIDS in Africa as a security threat to the United States' own interests, through its destabilizing of the structure of whole societies and economies.

This welcome recognition of the socio-economic impact of disease may help nudge the United States to take greater leadership and participation in global health issues. Multilateral initiatives are the way forward, but the experience with malaria has shown that the United States is reluctant to participate in ventures other than on its own terms. If Gore is serious, a first step would be for the United States to repay its debts to the UN technical agencies that are often in the front line of the war against disease in Africa and begin to play an active role in modernizing them and their programmes. ■





Laser ignition
'Unrealistic' NIF
project beset by
technical hurdles
p120



Superconductor
Bell Labs win prize
patent after a
12-year dispute
p121



Naples
Genetics institute
reverses trend by
moving south
p122



Sea squirts
Evo-devo sheds light
on our surprisingly
close relatives
p125

Celera turns to public genome data to speed up endgame...

Washington and Paris

Celera Genomics Corporation appears to have cut back on its plans to single-handedly complete a high-quality sequence of the human genome. The company now says that it intends to achieve the same end by combining lower-quality sequence with data from the international, publicly funded Human Genome Project (HGP).

At a press conference on Monday in Rockville, Maryland, Celera announced that it had itself sequenced 81 per cent of the human genome, and had combined this with publicly available data to cover 90 per cent of the genome.

Craig Venter, the Celera president, said the company plans to stop sequencing the human genome at the '4X' level in June — meaning that four bases of sequence have been generated for every base of genome — instead of 10X, as originally planned.

Celera's formation in 1998 led to an acceleration of the public project, which is funded by the US National Institutes of Health and Department of Energy, and Britain's Wellcome Trust (see *Nature* 395, 207, 1998). Its use of public data means that Celera will now sequence the human genome fewer times than previously planned. "If we had to only use our own data, we might have to go as high as 10X," Venter said.

Instead, it will combine these data with the 5X draft that the publicly funded project is expected to produce this spring. Venter described Celera's use of the public data as a "de facto collaboration", and said the company would "give attribution and credit" to the public project.

The acknowledgement seems to mark the end of Celera's original plan to 'shotgun sequence' the whole human genome on its own by sequencing millions of DNA fragments without knowing where they belong in the genome, and then use sophisticated software to work out which fragment goes where.

Now Celera will need to hang its sequence data on the framework produced by the public project. Celera used the same approach to sequence the *Drosophila* genome, in collaboration with the Berkeley Genome Sequencing Project (see *Nature* 401, 729; 1999).



Venter: modifying sequencing plans.

Celera claimed at its press conference to have sequenced 81 per cent of the human genome, but only to a depth of 1.75X. Combining these data with the public sequence has raised the total to 90 per cent, at 2X coverage.

Hamilton Smith, 1978 Nobel laureate in medicine and a co-founder of Celera, who is now its senior director of DNA resources, said the company's shotgun data match up well with the HGP's 'clone-by-clone' data. "Eighty per cent of the public sequence is covered in our data."

The public data represent about 50 per cent of the total in Celera's database, so the 20 per cent of unique data provided by the HGP represents a net 10 per cent contribution beyond Celera's, Smith explained.

John Sulston, director of Britain's Sanger Centre in Cambridge, where one third of the public sequencing is being carried out, points out that Celera's coverage is only to a depth of 1.5X, whereas the public project has

already completed half of the sequence at 5X.

He says the Celera press release could be rewritten to say that, with the company's 1.5 X data on top of the 5X data, "half [the genome] is really very well covered", and that Venter has helped "extend our 50 per cent".

Members of the public effort have agreed to deposit their sequence data every 24 hours. Because Celera hasn't accepted any public money, it is under no obligation to make its data public, but Venter repeated on Monday that it will eventually do so.

Venter estimated that the combined data contain 2.58 billion base pairs (Gb), and that 97 per cent of human genes are now in the company's database. But he declined to estimate the number of genes these data contain, adding that most genes are still in fragments.

Philip Green, a biocomputing expert at the University of Washington, estimates that the sequencing output of Celera and the public project are running neck and neck, at 1.5 Gb per month. He says the total amount of raw data required for adequate coverage of the genome could be obtained by the summer, provided the data from both groups are combined. "Neither group seems likely to get to that point by itself before the end of the year," he says.

Paul Smaglik and Declan Butler

...as Internet fervour hits genomics

London

When the dust settles over the recent dramatic upsurge in the value of biotechnology stocks — with the price of shares in some companies increasing three- or fourfold within a few weeks — a key role in triggering the goldrush could be credited to a US investment website visited regularly by many thousands of small investors.

The most dramatic news has come from Celera Genomics of Rockville, Maryland (see above). The company's announcement on Monday that it has compiled sequence data from its own and publicly funded efforts covering 90 per cent of the human genome sent its stock soaring, at one point

reaching a value of \$258, compared with \$186 the previous Friday.

Shares in other genomics and biotechnology companies received a parallel boost from the news, with the Nasdaq Biotech Index rising by 7.6 per cent. The value of shares in companies with a direct involvement in genomics has skyrocketed in recent weeks (see *Nature* 403, 4; 1999).

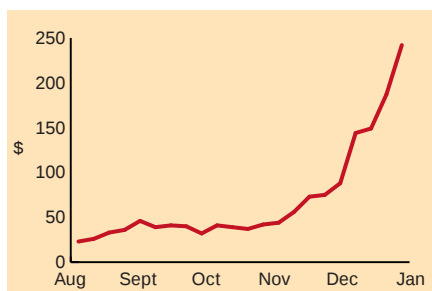
Stock analysts point to several recent announcements to help explain this surge of interest, including the public sequencing of the one-billionth human nucleotide pair (see *Nature* 402, 331; 1999) and the publication of the complete sequence of chromosome 22 (see *Nature* 402, 447; 1999).

Equally significant appears to have been an announcement nine days before Christmas by the managers of a website known irreverently as The Motley Fool (TMF) that they had decided to purchase \$50,000 of shares in Celera Genomics to add to their portfolio of 'rule breaker' companies.

The impact of the announcement, coming from a group of investors who had been among the first to spot key Internet stocks such as America OnLine — whose shares, purchased in 1994 at 46 cents, are now worth more than \$70 — was electric. While the number of shares purchased in Celera had previously been running at about 150,000 a day, this was multiplied overnight by a factor of more than ten.

The explosion in demand was rapidly reflected in the price of Celera stock. Having hovered around \$40 for much of the autumn, it had been growing steadily to reach a value of \$72 by 16 December, the day of the TMF announcement.

Within three trading days the value of the shares had shot up to \$125, and this momentum was maintained in the days to come. The shares reached a high for the year of \$193 on 30 December; the previous



On the up and up: Celera's share price has taken off dramatically in recent weeks.

August, they had been down to \$22.

After dropping slightly as a result of profit-taking in the days immediately after the Christmas break, the shares took off again, defying predictions that the pre-Christmas increase in value would be short-lived. Last Friday, trading in Celera stock was halted when the company announced plans for its Monday press conference.

Analysts say that many of those who have been buying up genomics stocks at such speed seem to be individuals who have already profited from the recent growth in the value of Internet stocks, and have been

attracted to genomics because of the similarities between the two fields.

"People have been saying that biotechnology is like the Internet, but until recently the stock has been much cheaper," says Genghis Lloyd-Davies of Credit Suisse First Boston.

Managers at TMF reflect this new-found enthusiasm. The decision to invest heavily in Celera was taken after assessing the extent to which the company fitted the strict criteria they have laid down for the selection of 'rule breaker' stock.

One of these, as described in a statement explaining the investment last month, was the fact that Celera appeared to be the dominant company in a rapidly growing area of commercial significance. "Celera has more 'momentum' — more speed — than any other company that we've ever bought," said the statement.

TMF managers acknowledge that their decision to invest in Celera reflects a long-term commitment. Dave Gardner, one of the founders of the website says: "We believe that in this coming century more value will be created in biotechnology than in any other technology."

David Dickson

Inadequate optics 'threat to US laser facility'

Washington

A series of technical obstacles could block completion of the US National Ignition Facility (NIF), according to an interim project review delivered this week to Bill Richardson, the Secretary of Energy.

The huge laser facility is under construction at the Lawrence Livermore National Laboratory in California. But major optical components may be unable to carry the NIF lasers at the intensity originally envisaged, says the review, by a task force of the Secretary of Energy's Advisory Board.

Optical damage will occur in these fused silica lenses at a lower intensity of laser light than was expected when the NIF was designed, because they will be operating in a vacuum — a conclusion that the authors admit is "a surprise".

They also say it will be hard to keep the equipment clean, because of "very cramped assembly areas" in the NIF building. And they warn that it will be difficult to order and procure 30,000 small optics components to adequate specification.

But the task force, chaired by John McTague, former technical director of the Ford Motor Company, concludes that the NIF can still be completed, although it will take longer and cost more than planned. "The task force has not uncovered any technical obstacle that would clearly prevent completion and operation of the NIF sys-

tem," said McTague, releasing the interim report on Monday (10 January).

The task force recommends that Livermore should assemble 96 lasers — half the original total — and use each of them at half their originally designed power to prevent damage to the fused silica optics. It then suggests a pause in the project before assembling the other lasers and raising the power intensity to the original design level by 2007 — four years later than originally planned.

Asked if overruns on the \$1.2 billion project would be much more than \$300 million — widely floated since problems emerged on the project last September — McTague said, "I don't know". He agrees that construction of the NIF began with "unrealistic" expectations, "because there wasn't enough fleshing out of the 'hows' of the project".

The task force criticizes Vic Reis, assistant secretary of energy responsible for defence programmes when NIF was started, for allowing it to proceed with contingency funding of only 15 per cent. This "may be appropriate for the construction of an apartment building, [but it] is too low for this challenging and unique project," says the review, adding that such projects should have a contingency of 30 or 35 per cent.

It also says that the "scientific culture" of the Lawrence Livermore laboratory hindered the execution of such a large engineering project, and criticizes its director, Bruce

Tarter, for failing to take direct responsibility for the NIF — even now. "The project needs to report directly to him," McTague says. "He needs to insert himself into the process." Tarter was unavailable for comment.

The task force found, however, that problems with cleanliness during laser assembly inside the NIF building could be solved by importing outside expertise. These problems were initially cited by laboratory officials as the main reason for the project's problems (see *Nature* **401**, 101; 1999), although the McTague review spells out far broader technical and management difficulties running across the entire project.

A spokesman for the Department of Energy said that it would respond to the review when the final report was delivered — it is expected in the spring. Critics of the NIF say the review's technical content shows that the project won't work. "What makes sense is to halt the project, build one beam and see if it works," says Chris Paine of the Natural Resources Defense Council, an anti-nuclear group that opposes the project.

But Burt Richter, former director of the Stanford Linear Accelerator Center and a member of the task force, rejects charges from environmentalists that the review had been too kind. "This is a hard-hitting report," Richter says. "It says that a laser intensity of half the design is a shoo-in, but anything else is quite risky."

Colin Macilwain

UK discussed ban on foreign job ads in 1960s

London

The British government was so worried about the loss of scientists and engineers to the United States in the late 1960s that it considered banning foreign recruitment advertising, according to documents released last week by the Public Record Office.

The documents, released under the 'thirty-year rule', also reveal proposals from a top civil servant in the Cabinet Office that scientists should be explicitly treated as an elite group in society, with benefits including special tax exemptions.

But reports from Germany, suffering a similar 'brain drain', showed that advertising bans had had little effect. The government was also worried that the press would react with hostility to such a suggestion.

Civil servants expected the Labour government of the time to be strongly opposed to making scientists a social elite. Even a merit award system for young scientists was scotched by the minister for education and science, Edward Short, as neither desirable nor necessary. Short warned that it could be a source of discontent and "future trouble" as it had been in the medical profession.

Officials at the time had difficulty in accurately assessing the extent of the 'brain drain'. Estimates of graduate emigration ranged from as little as 5 per cent to as much as 40 per cent in at least one university that was a producer of top-quality engineers.

In 1967 the United Kingdom had a net loss of about 2,300 active scientists and 3,600 engineers. There was evidence that people leaving the country were disproportionately well qualified.

Prime Minister Harold Wilson wrote to the minister for employment and productivity, Barbara Castle, stressing his concern at estimates that the United States would have a shortfall of about 400,000 engineers in the following decade, offering even more job opportunities to British graduates.

Cabinet Office officials considered the matter so important that they raised the possibility with the government's chief scientific adviser, zoologist Sir Solly Zuckerman, of treating scientists as "a preferred class of citizens". This could have involved preferential grants, reduced taxation and improved arrangements for professional mobility and pensions.

"Any proposals based on such conclusions would be politically impossible at present but if the statement [that it is vital to stop the brain drain] is correct, I believe we should face it honestly," says a letter to Zuckerman from F. H. Allen of the Cabinet Office. "There have been elite classes in the past: perhaps we shall deliberately have to create another in the future."

Zuckerman thought that improved uni-



Zuckerman (left) and Wilson: concerned at the lure of the United States for UK scientists.

versity administration would help, and advised vice-chancellors and principals to encourage moves towards a system like that of the United States. There, staff members were encouraged to accept external

consulting work and had flexibility in salary systems and in making research grant applications.

A document from the Ministry of Technology to the Cabinet Office suggests the reaction to a possible advertising ban was reflected in the stout defence by the science magazine *New Scientist* in the face of "an outcry" by readers to its acceptance of full-page advertisements from an American recruiting agency. The authors of the ministry paper point out that the magazine "stood firm on its right to inform scientists in this country of the opportunities elsewhere".

Zuckerman also advised the Foreign and Commonwealth Office to "limit" help requested by the Canadian government for Canadian firms trying to recruit British staff.

Natasha Loder

Bell Labs win superconductivity patent

Washington

Bell Laboratories, the research arm of New Jersey-based Lucent Technologies, has won a fierce competition for the US patent rights to yttrium barium copper oxide (YBCO). One of the most important high-temperature superconductor materials, this was identified by several rival groups of researchers in early 1987.

After twelve years of deliberation, a panel of four patent judges has ruled that Bell Labs can exercise the patents it filed on 3 March 1987, when at least four research groups in the United States were racing to discover oxide materials that would attain superconductivity at relatively high temperatures.

The race had been set off the previous year when Alex Müller and Georg Bednorz, two researchers at IBM research laboratories at Rüschlikon, Switzerland, discovered a copper oxide ceramic that could conduct electricity without resistance at a temperature of 35 Kelvin. Researchers then urgently looked for ceramics that would show the same property at more than 77 Kelvin — the temperature of liquid nitrogen, a cheap coolant that would allow widespread application of the materials.

Teams at IBM, the Naval Research Laboratory and the University of Houston filed patents for the successful compound soon after Bell Labs, which was then part of AT&T. Under US patent law, they were able to pursue an 'interference' procedure to establish which team had actually discovered the compound first, and was therefore entitled to patent it.



Coveted prize: polarized light micrograph of a YBCO high-temperature superconductor.

The finding of the US Patent and Trademark Office has not yet been publicly announced, but has been welcomed by scientists at Bell Labs. "The process has shown that we were the first to do it," says Bertram Batlogg, one of the team that made the discovery.

But some of their rivals continue to dispute that, claiming that Bell's rivals were merely unable to prove that they had made the discovery ahead of the filing date. "It was a tie, but they got the patent because they filed first," says one ex-IBM scientist who has followed the dispute closely.

The scientist also contends that the Bell Labs team "had no data" to prove that it actually had a material that was superconducting above 70 Kelvin

► throughout 90 per cent of its volume — the criterion the patent office said was needed to prove the value of the discovery. He contends that Bell's X-ray diffraction data, showing the crystalline structure of their material, were insufficient to prove 90 per cent bulk superconductivity.

Batlogg dismisses the distinction. "There's been a misunderstanding about us not having the data," he says. "The X-ray diffraction data showed that we had a single crystal phase, and the magnetic and transport data showed that we had a superconductor. I'm surprised people are

still so worked up about this."

Lucent has asked BTG, the UK-based technology licensing company, to sell licensing rights to other parties. One such agreement has already been reached with American Superconductor, a corporation based in Westborough, Massachusetts. Patents have also been issued to Bell Labs for YBCO superconductors in Europe since 1996, and in Japan earlier this year, but IBM also has patents in these countries for technology used to fabricate the materials.

It is not clear if IBM will appeal against the patent office's ruling.

Colin Macilwain

Genetics takes off in Naples as institute moves south

Milan

A major Italian genetic research institute will move south from Milan to Naples this summer, enhancing the city's reputation as one of the country's leading areas for genetic research and reversing the usual direction of scientific migration in Italy.

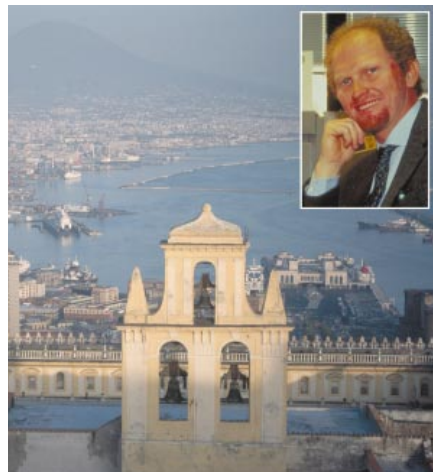
The relocation of the Telethon Institute of Genetics and Medicine (TIGEM) has been made possible with regional development funds from a variety of sources. These are intended to promote research and development in Italy's poorer southern regions.

TIGEM will move into an Integrated Centre for Biological Research that is being set up by the CNR, Italy's research council. The centre will also house the CNR's International Institute of Genetics and Biophysics (IIGB), best known for its pioneering work on bacterial phage genetics, and several smaller biological research institutes.

A further boost to Naples' reputation in genetics could come from a second initiative, also supported by development funds. Called Biogem, this 40 billion lire (US\$200 million) fund is intended to expand work on developmental biology, human genetics and cancer in Naples and its environs. Although approved in principle by the research ministry in 1997, the money is not expected to start flowing until spring this year.

The IIGB will be a focus for genetics in Naples. The centre will be located near Naples' major research hospitals and the Federico II University, in laboratories that were vacated by a pharmaceutical company in the mid-1970s and acquired by the CNR in 1996. European Union structural funds for underdeveloped regions have paid for refurbishing the building.

TIGEM was set up in 1994 by Telethon, a charity that raises money for research into genetic diseases through an annual television appeal. The organization's premises in



Promising outlook: Naples has the funds for two new research centres. Inset, Andrea Ballabio.

the San Raffaele Biomedical Science Park in Milan had become cramped, and last year it began looking for new premises for its staff of 75, including 40 scientists.

The CNR centre offers much more space and other scientific and financial incentives, says director Andrea Ballabio, himself a Neapolitan. Under an agreement with the research ministry, the CNR will give TIGEM laboratory space in its new centre rent-free, while the 5 billion lire costs of the move are coming out of ministry funds allocated to support research in the south.

The Campagna region, in which Naples lies, will provide further funding for running TIGEM during its first three years in Naples. Telethon, which will continue to provide TIGEM's annual 4.3 billion lire running costs, does not intend to increase the number of Telethon staff, despite the extra space available. "We hope to have space to host scientists from other research institutions such as CNR or the universities," says Ballabio.

TIGEM scientists already work closely

with scientists in Naples, which has Italy's second-highest concentration of human geneticists after Milan. Such collaborations resulted in 1997 in the identification of the genes for spastic paraplegia, and, last year, for lysinuric protein intolerance — a rare disorder causing growth delay and sometimes mental retardation. Future work will attempt to identify genes involved in retinal disorders, among others. When TIGEM and the IIGB merge their gene-sequencing facilities, Naples will probably have the largest sequencing capacity in Italy, says IIGB director John Guardiola.

Guardiola wants to work with TIGEM scientists on screening the genetically homogeneous populations of local villages, where intermarriage is common, to try to identify carriers of rare genetic diseases.

Modern Italian genetics began in Naples. Adriano Buzzati Traverso was the IIGB's founding director in 1962, and Luigi Luca Cavalli-Sforza was director in the mid-1960s. It was there that the first gene on the X chromosome was identified in 1981. The university also has a strong genetics programme.

Because not all TIGEM scientists will make the move, the prospect of vacancies for prestigious Telethon positions is delighting local geneticists. Guardiola says that many geneticists who trained in Naples and subsequently left have already expressed interest in applying for the jobs. "It is an opportunity to come home to an even stronger scientific environment than the one they left — and with the most modern facilities," he says.

Ballabio hopes that TIGEM researchers will benefit from the Biogem initiative. Half of Biogem's grant will support projects in developmental biology, human genetics and cancer, and the expansion of research services, including bioinformatics, sequencing and the development of mouse models of disease, in various research institutions.

Roberto Di Lauro, a professor of human genetics at the Federico II University of Naples, and secretary of Biogem's scientific and technological council, says that Telethon and Biogem intend to work closely together to avoid duplicating efforts.

Thus, sequencing will only be done at the new CNR centre, he says. Genomics — mostly the development of mouse models of human diseases — will be expanded at an institute to be built with the remaining 20 billion lire of Biogem money at Ariano Irpino. This small town, in the mountains some 100 kilometres from Naples, is the home town of research minister Ortensio Zecchino.

The remote location of this research institute has been criticized by many Italian scientists. But Di Lauro counters that "no one in Naples came forward with a site in the city", and that "while critical mass is important, it is also important to let research start up in new locations".

Alison Abbott

NIH seeks funding for neuroscience centre ...

Washington

Officials at the US National Institutes of Health (NIH) are hoping that President Bill Clinton's 2001 budget request will include funding for work to begin on a \$270-million neuroscience research centre.

Construction of the centre, to be located on the NIH campus in Bethesda, Maryland, would start with a 20,000 square metre building costing \$70–90 million. This would be followed by additions and renovations to an existing building.

Gerald Fischbach, director of the National Institute of Neurological Disorders and Stroke, and Steven Hyman, director of the National Institute of Mental Health, have been discussing bringing together NIH intramural neuroscience research into a single building over the past year.

Only recently has the need for a new facility been agreed, however. Former NIH director Harold Varmus expressed his support for the project at his final advisory committee meeting last month, and the White House is expected to be sympathetic to the request.

Fischbach and Hyman decline to say how much money they are seeking for the project's first year until Clinton's 2001 budget request is released next month. Congressional approval would be needed for the construction project, which is likely to take about eight years to complete.

One reason Fischbach and Hyman want to build the centre is to help reverse the fragmentation that has occurred as neuroscience and brain research have evolved at various NIH institutes. "You don't want the research to suffer because it's been fragmented," says Fischbach.

He points out, for example, that changes in the activity of neurotransmitters such as dopamine are likely to contribute to schizophrenia, Parkinson's disease, drug addiction and chronic pain — each of which is currently studied at a different institute.

After neuroscientists are moved from nine separate institutes (including two that are off-campus) into a single complex, research should be reorganized by themes to promote projects ranging from basic to clinical research, Fischbach says. He cites neurodegeneration, neurogenetics and pain as possible themes.

Hyman agrees that combining various approaches will help bridge basic and clinical science. Creating the complex near the NIH's Clinical Research Centre, a \$333 million building scheduled for completion in 2002, should further that goal, he adds.

"It makes sense to physically unify the research on campus, in order to have proximity to the critical mass of molecular biology and genetics research on the one hand,

and to be near the new clinical research centre on the other," Hyman says.

The new building will also be highly flexible, Fischbach and Hyman say, with moveable walls, lab benches that can be adjusted for both *in vitro* and animal work, and space for imaging equipment that scientists from the multiple institutes can share. Sharing research equipment and reducing duplication should cut costs, they add.

Dennis Choi, president of the Society for Neuroscience and a neurologist at the Washington University school of medicine in St Louis, Missouri, says he is "enthusiastic" about the proposal and agrees that it could promote greater efficiency. The complex should bring together basic and clinical neuroscience, he says: "There's a lot happening at that intersection."

Beyond the arguments of unity, economy and synergy, Joseph Coyle, chairman of psychiatry at Harvard Medical School, says the project can be justified by the direction in which neuroscience appears to be heading.



Once the human genome has been sequenced, "the real focus is going to be on how the genome constructs the brain," says Coyle, pointing out that more than half the human genome is likely to be related to the brain or nervous system. "Scientifically, there really is a good reason for getting the centre together."

Paul Smaglik

... as lobbyists anticipate more money for all

Washington

US President Bill Clinton is expected to seek substantial increases in funding for basic scientific research in his budget request for the 2001 financial year, to be released on 7 February.

Science lobbyists say the request may include \$1 billion (6 per cent) more for the National Institutes of Health (NIH) and a "double-digit increase" (at least \$350 million) for the National Science Foundation (NSF) — the largest increase that Clinton has requested for the agency in his eight years in office.

Clinton is expected to announce a broad effort to boost spending on scientific research in his State of the Union address on 27 January. Other important science-funding agencies, including the space agency NASA, the Department of Energy and the Department of Defense, are also expected to benefit.

Some of the new money will be allocated to targeted initiatives in computer research, nanotechnology and environmental science. But the president also intends to propose substantial increases in funding

for basic research in most scientific disciplines.

Unlike similar increases he proposed two years ago, these are not expected to be conditional on a tobacco tax or any similar budgetary gimmick. However, they will exceed the spending caps agreed by Congress and the White House in 1997. The caps were massively breached during last year's budget negotiations, but have not yet been formally abandoned.

Clinton will propose substantial additional funding for the NSF to continue a research initiative in information technology begun last year, to lead an interagency nanotechnology initiative, which will be unveiled in the budget (see *Nature* 400, 95; 1999), and to start a build-up of research in the environmental sciences that was proposed last autumn by the National Science Board (see *Nature* 400, 492; 1999).

Michael Lubell, head of public affairs at the American Physical Society, welcomed reports of the budget proposal, saying it reflects recent calls by a coalition of scientific societies for a

"balanced portfolio" of research spending.

But John Porter (Republican, Illinois), chair of the appropriations subcommittee in the House of Representatives that funds the NIH, said he was "disappointed" with the president's reported plan to propose a \$1 billion increase for the NIH. Porter points out that it "falls short" of a hoped-for 15 per cent increase, which would double the agency's budget in five years. He adds that the president failed to support doubling NIH funds during early stages of the budget process in the past two years. Congress gave the agency increases of 15 per cent or more in 1999 and 2000, which the president ultimately approved.

Ray Merenstein, vice-president of Research!America, a lobby group for biomedical research, agrees that the proposed increase for the NIH is "not high enough". But he is optimistic about the outcome, as the president's initial offer is still one of the highest percentage increases Clinton has proposed for the agency since taking office.

Colin Macilwain & Paul Smaglik

Greenwich 'could be symbolic timekeeper for electronic era'

London

Britain is bidding to make the Royal Observatory at Greenwich the symbolic timekeeper of the Internet by launching Greenwich Electronic Time (GeT).

The bid is being led by a consortium of more than 4,000 European online retailers from 20 countries, known as the Interactive Media in Retail Group (IMRG). IMRG hopes that GeT will become the standard time for all electronic commerce.

Consortium officials also say that GeT would smooth the path of electronic transactions on the Internet; at present, they argue, a lack of standardization in e-commerce means that goods can be owned by two parties at the same time.

Despite its name, GeT is not Greenwich Mean Time: it is another way of describing Co-ordinated Universal Time (UTC). Based on an agreement by more than 200 countries with atomic timekeeping, UTC uses an average of the time as measured by atomic clocks in these countries. It is so accurate that it occasionally needs to be adjusted with 'leap seconds' to match the slightly irregular rotation of the Earth.

UTC is already used worldwide, and is popular among the commercial and scientific sectors in the US. But although available on the Internet, its existence is not widely recognized.

"For me the benefit of the GeT project is the publicity and marketing, establishing that this is the time standard on which we need to operate on the Internet," says John Lavery, head of the time section at the UK National Physical Laboratory in Teddington. He says there is "a lot of ignorance about what time standards are", and hopes GeT will raise awareness of time "etiquette".

Lavery has been advising IMRG about the new project. He says one of the reasons he is open to the idea is because of Swiss watchmaker Swatch's recent introduction of Biel Mean Time, which measures a day as 1,000 'beats' and created a new meridian in Switzerland.

"They've taken the start of the universal day as Central European Time," says Lavery. "This is convenient for Swatch headquarters, but if you are somewhere else in the world, dealing with Global Positioning System and UTC technology and your day starts at a different time, this is crazy."

Details of the proposal for GeT can be found at: www.get-time.org. **Natasha Loder**

Brussels research chief backs European website proposal

Paris

European Union research commissioner Philippe Busquin is backing plans by the European Molecular Biology Organization (EMBO) to establish a website called 'E-Biosci'. This will be a European counterpart to PubMed Central, the free website for life-science papers to be launched this month by the US National Institutes of Health.

E-Biosci has Busquin's "full support", he told *Nature* this week. "It may have important implications, not only for the way scientific information is spread and accessed worldwide, but also for the way business will be run in this sector in the near future," he said.

An official from the European Commission's research directorate says the EC is open to paying part of the scheme's estimated US\$3 million annual running costs. Although EMBO has agreed to pay 500,000 Euros (\$511,000) to start E-Biosci, the project has no firm plans as yet for long-term funding.

The commission will be represented next week at a meeting convened by EMBO in a bid to take E-Biosci forward. Representatives of European research councils and the Wellcome Trust will also attend, along with major publishers including *Nature*, Oxford University Press and Elsevier Science.



Busquin: encouraging E-BioSci partnership.

EMBO has until now been the driving force behind E-Biosci. The meeting aims to promote wider involvement, to reach agreement on the structure of E-Biosci and to seek commitments from public research funders.

"The project's success will require a true partnership between all the actors concerned and I would be ready to give my full backing to such a partnership," says Busquin.

The commission official says he hopes the meeting will result in the creation of a consortium to finalize and run E-Biosci, adding that such a consortium would be eligible for EU funding. "It is important that we have a credible alternative to PubMed Central, as without this we will be in a weak position to negotiate with the United States."

The sum required to run E-Biosci is relatively small, he points out, and the commission is allowed to fund such ventures under 'accompanying measures' to the Framework research programme.

Declan Butler

Japanese support biotech start-ups

Tokyo

The Japan Development Bank last week announced that it is to set up a dedicated biotechnology fund. The move reflects the growing willingness of Japanese investors to nurture domestic start-up companies.

Biotechnology-related investments by Japanese venture capitalists have tended to focus on established companies or foreign start-ups. In contrast, the new BioIncubation Fund will target domestic companies early in their development.

More Japanese scientists want to become entrepreneurs, following the announcement of new regulations for technology transfer from universities and government laboratories. There has also been a surge in public support for technology-based businesses.

About 20 biotechnology companies are believed to have been set up in Japan during the past year. But Japanese venture capitalists, many linked to large banks or security houses, have been slow to invest in start-ups.

"The increase in the number of newly established companies is very significant," says Robert Kneller, a professor in the

department of intellectual property at Tokyo University's Research Center for Advanced Science and Technology.

Akiko Itai left the University of Tokyo five years ago to set up the Institute of Medicinal Molecule Design, a company developing software tools for drug design. "Many young graduates still prefer the security of a position with a large pharmaceutical company or a national university to working on an annual contract for a venture business," he says. Recruiting researchers from large companies is not an option, says Itai, because venture businesses can rarely match their salaries.

Biotechnology is not the only field to profit from the surge in venture capital. With a Japanese branch of the Nasdaq shares index opening its doors next summer, large companies and financial institutions are rushing to invest in technology-based companies.

Only about a third of the biotech companies set up last year originated in academia or government laboratories. This suggests that, in Japan, corporations may be a better source of entrepreneurial talent than universities for some time to come.

Robert Triendl

Biologists flock to 'evo-devo' in a quest to read the recipes of life

How has the diversity of animal and plant forms evolved? Researchers are combining evolution, development and genetics to address this hot topic.

Atlanta, Georgia

The emerging discipline of evolutionary developmental biology — nicknamed 'evo-devo' — was the star turn at last week's annual meeting of the Society for Integrative and Comparative Biology, held in Atlanta, Georgia.

Two symposia, sponsored by the US National Science Foundation, were devoted to the discipline, and addressed both animal and plant research. The society, formerly known as the American Society of Zoology, also created an evo-devo division. In recent months, two journals focusing on evo-devo research have been formed.

The meeting of the 2,200-member society drew more than 1,200 scientists, the most in recent years. This is attributed largely to the evo-devo sessions, which attracted European scientists.

One of the symposia focused on the 'Hox' gene clusters. These are believed to be integral to the structuring of animal bodies during development, and to the evolution of morphology (the shape and form of organisms). The second symposium linked the history of evo-devo to methods for combining techniques from molecular biology with older disciplines such as phylogeny (study of the evolutionary relationships between species) and morphology.

Hox of delights

Evo-devo is described by symposia organizer Billie Swalla, a zoologist at the University of Washington, Seattle, as the way in which "changes in developmental genes cause evolution and morphology". Zoologists believe that they are looking at "profound developmental genes" that provide the signals responsible for creating various parts of the organism.

As an example, she cites her research on the development of marine animals called tunicates (sea squirts), from their distinctive larvae. While the fossil record for tunicates is negligible, possibly because they have no hard parts, she says that the use of studies of genes and morphology to reconstruct their evolutionary relationships has shown that tunicate larvae have repeatedly gained and lost their tails during the course of evolution: a small cassette of genes seems to be responsible for the switch.

These questions are the stuff of evo-devo,



Mighty squirts: tunicates are among the closest living invertebrate relatives of the vertebrates. Study of the genes that control their development can help to reveal how this evolutionary transition occurred.

she says. Tunicates are also interesting because they are closely related to vertebrates. During the evo-devo symposia, scientists discussed studies of animals ranging from butterflies to the chick.

Corresponding Hox gene clusters instruct cells to produce organism parts in humans, chicks, flies and other animals. Sean Carroll of the University of Wisconsin at Madison presented studies showing how a regulatory gene called *hedgehog*, identified in the fruitfly *Drosophila melanogaster*, directs the formation of eyespots on the wings of butterflies (see *Nature* 384, 236–242; 1996).

To ensure a broad approach, and to draw analogies between evo-devo studies of plants and animals, the society included plant systematist Michael Donoghue, director of Harvard University's herbaria.

Scientists have found Hox genes in plants, but they are not thought to be as deeply involved in the organism's body plan as they are in animals. Michael Purugganan, of North Carolina State University, reported that another class of genes, called MADS-box genes, control plant development in ways analogous to the Hox genes of animals (see *Nature* 399, 144–148; 1999).

Although this line of work is well known

to botanists, few animal biologists are aware of it. It was felt that an approach combining work on plants and animals would yield more of the comparative information on which evo-devo is increasingly based.

Possibilities and constraints

Rudolph Raff, a zoologist at the University of Indiana and the first president of the society's evo-devo division, says that he expects the discipline to answer how the course of development varies between organisms, how this evolves, and how the evolutionary 'choices' made during development constrain evolution.

Developmental constraints are one reason why organisms cannot evolve into an infinite variety of forms. Evo-devo seeks to find out why, and what these constraints are.

A number of participants at last week's meeting said that, as the field matures, they expect researchers to study the genomes of many species — not just standard models such as the nematode worm or mouse.

"Molecular technologies are permeating all disciplines," says Martin Feder, a zoologist at the University of Chicago who is president of the society and a strong supporter of the evo-devo symposia. "Evolutionary studies are now permeating all disciplines. We stand at the crossroads of this."

But some researchers, such as Paula Mabee, a specialist in phylogenetic systematics at the University of South Dakota, warn that evo-devo still needs the foot soldiers of basic zoology.

Raff agrees. "We could fool ourselves pretty badly" with an over-reliance on genetic tools, he says.

Rex Dalton

Evo-devo seeks to answer how development varies between organisms.

European group opposes plans to axe French synchrotron

Paris The European Round Table for Synchrotron Radiation and Free Electron Lasers is opposing the French government's plan to close its own facilities in favour of a UK project.

The group, representing scientists who rely on synchrotron X-rays for their research, has written to prime minister Lionel Jospin opposing plans to close national synchrotron facilities at the Laboratoire pour l'Utilisation du Rayonnement Electromagnétique (LURE), near Paris, which is scheduled to close within five years. The government intends to cooperate with British plans to build a 3-GeV machine called Diamond.

"Thanks to the renown of LURE, France is a world player in a field essential to scientific and technological research," the letter states. "Unfortunately, the current strategy risks destroying an exceptional national legacy, condemning French researchers to a background role, and France itself to isolation."

Taiwan extends tax credits to boost industry R&D

Tokyo The Taiwanese government is set to approve a 10-year extension of a

comprehensive bill aiming at the "upgrading of industries". In its revised form, which is to be formally approved later this month, the bill will provide tax credits of up to 25 per cent for investments in research and development (R&D) and training.

All industrial sectors have previously benefited alike from tax credits for R&D and training, but the revised bill is likely to limit such credits to sectors such as semiconductors, personal computers and telecommunications equipment. Critics of the bill have pointed out that these are areas in which Taiwanese industry is already competitive.

India promises more money for research

New Delhi The new millennium has brought some good news for the fund-starved Indian scientific community. Addressing an audience of nearly 4,000 researchers at the Indian Science Congress in Pune last week, prime minister Atal Behari Vajpayee promised greater funding for research and development and vowed to remove "restrictive bureaucratic hurdles".

"India's investments in research and development are wholly inadequate and sub-critical," Vajpayee admitted, adding that the country was "determined not only

to be a participant in the information technology revolution but also to be in its vanguard". Noting that India's successes in information technology are already well known, he asked the Indian scientists to replicate them "in all areas of science and technology".

Climate coalition rocked as another car-maker quits

Washington DaimlerChrysler, the international motor manufacturer created last year by the merger of Daimler Benz and Chrysler, has withdrawn from the Global Climate Coalition (GCC), the main industry group lobbying against mandatory action to cut greenhouse gas emissions in the United States.

The withdrawal of DaimlerChrysler, following that of Ford late last year, leaves General Motors as the only major car manufacturer still supporting the GCC's position on global warming.

In a statement, the GCC said that DaimlerChrysler's decision was "disappointing, but not unexpected", adding that both the car-maker and the GCC would continue to oppose the Kyoto protocol while "pursuing an aggressive policy of voluntary actions" to address concern about emissions.

California plans three campus-based centres

San Diego The governor of California, Gray Davis, proposed last week that the state should spend \$300 million over the next four years to create Institutes of Science and Innovation at three University of California campuses.

The proposal, which has yet to be approved by the state legislature, calls for a competitive review process involving all nine University of California campuses, with the winners each receiving \$75 million to build and equip the institutes. Each winning campus would engage business partners, who would jointly provide at least \$25 million annually towards each institute's operating budget.

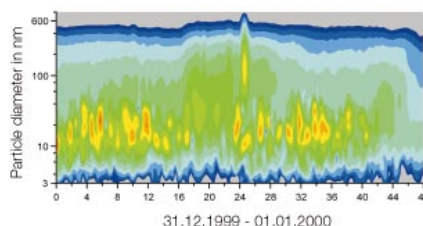
WHO pushes ahead with drive to eradicate polio

London The World Health Organization (WHO) this week launched its final push towards eradicating polio by the end of 2000. The campaign began in 1988 when there were 350,000 estimated cases of the disease. Last year 5,200 cases were reported.

The cost of wiping out polio is estimated at \$1 billion, although the campaign's funds are currently \$300 million short. But the WHO

says that eradicating the disease and ending immunization will save the world \$1.5 billion per year. "We are on the verge of an historic public health victory," said Gro Harlem Brundtland, director-general of the WHO, urging 30 heads of state in sub-Saharan Africa and South Asia to work to eliminate reserves of poliovirus.

Germany celebrates a dusty new millennium



Leipzig This dramatic but short-lived cloud of dust particles occurred over Leipzig shortly after midnight on 1 January 2000 — coinciding with celebrations of the new millennium. It was recorded by the particle spectrometer of the Institute for Tropospheric Research, which monitors the urban aerosol over the city in eastern Germany. Each particle had a diameter of about 100 nanometres.

Berlin boosts research funding despite squeeze

Munich The economic minister for Berlin, Werner Müller, is to devote a higher proportion of his shrinking budget to support technology transfer and innovation in the city, according to a report in the *Süddeutsche Zeitung*. Despite the general budget squeeze in Berlin, Müller will raise his support for research and development from its current annual level of DM1.7 billion (\$890 million) to DM2 billion by 2002.

According to the newspaper, Müller says his main aims are to bring venture capital into contact with high-tech research, and to bring small and medium-sized enterprises into contact with academic researchers.

UK science museum appoints new head

London Lindsay Sharp, president and chief executive of the Royal Ontario Museum in Canada, has been appointed as the next director of the Science Museum in London.

Sharp has been at the Royal Ontario Museum for the past three years. Between 1978 and 1988 he was the founding director of the Powerhouse Museum in Sydney. He will take up his appointment in July.

Global bodies won't save the environment: it needs grass-roots efforts ...

Sir — James, Gaston and Balmford in their Commentary provide confused arguments for a global conservation strategy¹. The fault lies in their jump from an exercise in numerical estimation to a prescription for global economic policy. Their proposal — for global institutions to buy land in less-industrialized countries and turn it into nature reserves — overlooks the enormously dangerous mid- and long-term consequences it would have, both on the environment and on the people who live there.

The unrealistically low prices paid for land in developing countries would not provide the necessary economic basis for “the losers” to find “alternative employment opportunities”. Most of those “losers” (cultures and communities that have maintained biological resources over thousands of years^{2,3}) do not consider their relationship to those resources as a job to be given up for money. Besides, we need only look at the failure of retraining and employment programmes in the United States to gain a sobering insight into the realities of such a proposition.

It seems more likely that the low-cost social programme James *et al.* suggest to compensate these people would increase the flood of migration to infra-urban misery belts around cities in the Third World, rather than causing a utopian transformation of rural and forest communities. Numerous studies have linked this migratory pattern to the loss of biological diversity in depopulated rural areas⁴.

James *et al.* dismiss community-based initiatives as not “encouraging... despite a decade of practice”. They favour instead a top-down programme directed by global institutions and carried out by national governments. Yet this approach, widely tested for several decades, has a dismal environmental track-record.

By contrast, there are numerous success stories of sustainable biological resource use grounded on community-based initiatives^{4,5}. Many of these have disappeared in recent years, but are now being rediscovered by biologists and anthropologists^{6–7}. Others still exist, but because they are adapted to their own sites they are small and hardly noticeable in a global analysis. However, their mere survival in the face of ineffective trickle-down approaches such as that advocated by James *et al.* testifies to their resilience and potential⁴. Indeed, given the spectacular failure of globalized, one-fits-all proposals during the past six decades, I would argue that any possibility

of developing an ecologically sustainable future will originate in small, community-based initiatives.

Many successful, small-scale purchase-for-protection programmes are limited to areas with relatively low cultural diversity and/or low levels of human interaction with diverse biological resources. But even here, buying a piece of land does not guarantee the conservation of the resources associated with it, especially in remote areas inhabited by marginal biological resources and human communities.

“Preserved” areas are damaged not only by chain-saws, but also by the introduction of alien invasive species, the accumulation of fuel in fire-protected areas, and the demise of human-adapted species and ecosystems. Human communities can protect biological diversity against these threats, using fine-scale management. This cannot be done by remote control from the boardrooms of the World Bank^{3–5,8}.

Ignacio H. Chapela

Department of Environmental Science, Policy and Management, University of California, Berkeley, California 94720, USA

1. James, A. N., Gaston, K. J. & Balmford, A. *Nature* **401**, 323–324 (1999).
2. Fairhead, J. & Leach, M. *Africa* **66**, 14–37 (1996).
3. Gomez-Pompa, A. in *Rainforest Regeneration and Management* (eds Gomez-Pompa, A., Whitmore, T. C. & Hadley, M.) 3–18 (Parthenon, New Jersey, 1992).
4. Berkes, F. & Folke, C. in *Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience* (eds Berkes, F., Folke, C. & Colding, J.) 1–125 (Cambridge Univ. Press, 1998).
5. Getz, W. M. *et al. Science* **283**, 1855 (1999).
6. Gomez-Pompa, A. & Kaus, A. *Proc. Natl Acad. Sci* **96**, 5982–5986 (1999).
7. Fairhead, J. & Leach, M. in *The Lie of the Land: Challenging Received Wisdom on the African Environment* (eds Leach, M. & Mearns, R.) 105–121 (International African Institute, Oxford, 1996).
8. Janzen, D. *Proc. Natl Acad. Sci* **96**, 5987–5994 (1999).

... or should private enterprise take over?

Sir — James *et al.*¹ call for the elimination of government subsidies that encourage ecologically harmful activities. This goal is admirable because it would remove government incentives that lead to resource abuse and would lessen the tax burden on the citizenry. However, proponents of this idea should be wary about redirecting the subsidies to protection of biodiversity.

Using subsidies to set aside a significant portion of a nation's land base in nature reserves can still have serious economic consequences if it means that other economic activities such as oil and gas development, timber production, or ranching are disallowed. Already suffering from impoverished conditions, Third World countries will not readily embrace the idea of reduced economic outputs. In addition, such a strategy ignores the possibility of

creating a profitable land base of enhanced biodiversity, and may also intensify development pressures on lands not included in the nature reserve system. As a consequence, species residing on these lands are likely to suffer a loss of habitat².

Creating a land base that eliminates options for other economic activities does not necessarily protect species. Newmark found that national parks in western North America have lost 42 populations of mammals and are in danger of losing more³. Also in national parks, Wagner *et al.* found declines in “genetic diversity within species populations, species diversity, ecosystem diversity, and habitat of landscape diversity”⁴.

There are better options for enhanced biodiversity, without government spending. Non-government enterprises that protect biodiversity include the McIlhenny family's Avery Island and the National Audubon Society's Rainey Preserve in Louisiana, the Fort Apache Reservation in Arizona, the Joint Venture Partners' preserve along the Kikori River in Papua New Guinea, and those run by Conservation Corporation Africa. They prevent trespass and generate funds from limited commodity development or from the wildlife itself. Such private options should be explored.

Matthew Brown, Donald R. Leal

Political Economy Research Center, 502 S. 19th Avenue, Suite 211, Bozeman, Montana 59718, USA

1. James, A. N., Gaston, K. J. & Balmford, A. *Nature* **401**, 323–324 (1999).
2. Krebs, J. R. *et al. Nature* **400**, 611–612 (1999).
3. Newmark, W. D. *Nature* **325**, 430–432 (1987).
4. Wagner, F. *et al. Wildlife Policies in the US National Parks* (Island, Washington DC, 1995).

Proteins and the naked truth about e-commerce

Sir — Vani Kalyanaraman (*Nature* **402**, 118; 1999) objects to the photo in the advertisement on the back cover of the 7 October issue of *Nature*.

Perceptions differ. Far from being “unpleasant”, the picture in my opinion is sensuous and certainly not pornographic, not even erotic. Together with the text, “The beauty of protein folding”, it made me sufficiently curious to read the entire advertisement.

The contortions through which the human body can go depend on the bending properties of muscles, tissue, sinews, ligaments and whatnot, and these must in turn depend upon the properties of proteins.

On page xvii in that same issue there is an advertisement headed “Human tissues” with a picture of a naked body that shows more explicitly what humans are like in the nude than the photograph in question, but Kalyanaraman does not comment on it. I think *Nature* has exercised its editorial

prerogative wisely in both cases and should guard against undue censorship.

Dattatray Parasnis

Gnejsstigen 6, 977 53 Luleå, Sweden

Sir — Vani Kalyanaraman objects to nudity in advertisements for scientific products, but I believe that nakedness might be more appropriate here than Kalyanaraman realizes (*Nature* **402**, 118; 1999). Given that we are being increasingly encouraged to buy almost everything via the web, perhaps use of nudity reminds us of the strong association between images of nakedness and e-commerce?

Alex May

Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

EPA error risk halving India's rice harvest

Sir — Methane has attracted particular attention as a greenhouse gas, as once it is formed, mitigation strategies are not easy to devise. The US Environmental Protection Agency (EPA) estimated in 1990 that global methane emission is about 550 Tg (550 million tonnes) per annum, with about 110 Tg of this released from paddy cultivation. This estimate turned out to be wrong, but it could have had devastating consequences for the economy of India.

Nearly 90% of the world's rice production comes from Asia, and rice is the staple food for 2 billion people. India produces about 80 million tonnes from about 42 million hectares. The EPA estimated methane emission from India as 37.8 Tg per annum in 1990¹. The implications of this estimate were so serious that India's National Physical Laboratory led an international collaboration of 15 laboratories to measure methane emissions directly².

Methane released during paddy cultivation was collected and analysed, with careful controls. Methane fluxes were typically between -10 and $+80$ mg per hour per m². Samples were taken from 34 sites in India of varying agricultural conditions, and measurements were done at different cultivation stages, from seedlings to harvesting, and integrated over a full year to cover single and multiple crops. Methane emission depends on the paddy-water ecosystem, and this was also measured in various ways over time.

All this direct agricultural information was used to calculate the methane emission budget³, which was 4.07 ± 1.25 Tg per year, only one-tenth of the 1990 EPA estimate. A repetition of the measurements in 1996 gave essentially the same result.

The main reason for the difference between the EPA estimates and the actual

measurements is that long incubation periods lead to anaerobic conditions, favouring high methane genesis. The EPA extrapolated the 1990 data without realizing that the high methane flux conditions had resulted in an annual methane emission from India that was an order of magnitude larger than the experimentally measured values.

This error has now been corrected and the estimate of global methane emission from paddy fields has been scaled down to 60 million tonnes. Methane emission from rice cultivation in India and other similar Asian countries is no longer considered a major factor in global warming.

If this error had gone uncorrected, international protocols would have required Indian methane emissions to be brought down to global average levels, reducing paddy cultivation by at least 47% in the short term. The economic cost of this would have been about 135,000 million rupees (US\$3.1 billion) per annum, nearly double the annual budget of all India's science and technology ministries combined³.

So the investments that enabled the National Physical Laboratory to perform and coordinate precise measurements saved the country from an economic crisis. This story illustrates that, even in developing countries, it is wise for governments to invest in research and development.

E. S. R. Gopal

Department of Physics, Indian Institute of Science, Bangalore 560 012, India

1. Ahuja, D. R. *Estimating Regional Anthropogenic Emissions of Greenhouse Gases* (US EPA Technical Series, Washington DC, 1990).
2. Parashar, D. C. *et al.* *Chemosphere* **33**, 737–757 (1996).
3. Gopal, E. S. R. *et al.* in *Proc. 2nd Int. Conf. Metrology, Quality and Global Trade* (eds Banerjee, P. *et al.*) 432–439 (Alpha, New Delhi, 1999).

Wherever HIV originated, polio vaccine is safe now

Sir — John Moore's review of E. Hooper's *The River* (*Nature* **401**, 325–6; 1999), on the possible link between the origin of HIV and a contaminated polio vaccine used during a mass vaccination campaign in equatorial Africa in the 1950s, carefully discussed the weakness of the book's hypothesis. But it is important to add that the polio vaccine currently being used worldwide is safe.

This message was not very clear when newspapers in Uganda first picked up the HIV–polio theory from the US media some years ago. As a result, some people thought that their child would be given a contaminated vaccine. So they kept their children from being vaccinated, putting them at high risk of poliomyelitis. In 1999, there were 49 confirmed polio cases in Uganda, according to the World Health Organization.

With the recent renewal of interest in the HIV–polio theory, there is a need to stress

the safety of the present polio vaccine whenever it is discussed. Hopes for global eradication of polio depend largely on vaccination coverage in all parts of the world.

Ka-Wing Wong

Department of Molecular Biology & Microbiology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, Massachusetts 02111, USA

East German academics faced unfair hurdles

Sir — We have some doubts about statements attributed to Jens Reich in a recent Briefing (*Nature* **401**, 637; 1999), that some evaluation interviews conducted by the Wissenschaftsrat in Germany were held in English. Although some evaluation committees may have interviewed an East German lecturer in English or may have required someone to give a lecture in English, the claim that interviews were conducted in English must be due to a misunderstanding.

East Germany's academics were not capable of conversing in English at this level. Real proficiency in English would have aroused suspicion — either of a rightist (non-socialist) leaning or, even worse, of collaborating with the West. How could scholars or scientists have learnt a foreign language in a country such as East Germany, which disadvantaged classical and humanist studies and allowed the modern language teaching of Russian to dominate?

Of course, a knowledge of English in the former East Germany was not unheard of. A few academics who travelled abroad endeavoured to learn the language. At the end of the GDR period, language training centres were established to provide quality teaching in Russian, English, French, Spanish and Portuguese, following the example of UK language schools. These centres gave academic and other specialist staff chosen for foreign assignments an opportunity to learn their target language.

The Briefing may have overlooked an important angle. Immediately after the fall of the Berlin Wall, many heads of institutions and departments in their fifties were replaced by members of the 'pretender' generation — biting the hand that had once fed them. If reunification hadn't happened, these 'pretenders' would only now be starting to reach those positions — but they would have been properly prepared by their mentors.

So, while we needed people who would stand by their principles, it was once again the Fouché-like opportunists who — believing a political fault is worse than a crime — left the field as victors of the game.

Bernd Legler

Erfenschlager Str. 150, D-09125 Chemnitz, Germany

Guy Moore

Hauptstr. 73, D-53619 Rheinbreitbach, Germany

Sowing nuclear misconceptions

In its recent deliberations over the Comprehensive Test Ban Treaty, the US Senate was not well served by the directors of the US weapons laboratories.

Kurt Gottfried

The US Senate's vote against ratification of the Comprehensive Test Ban Treaty (CTBT) last October casts a dark shadow over the treaty's prospects, and poses a serious threat to the global non-proliferation regime. Although bitter partisan politics and the Clinton administration's inept handling of the case for the treaty were dominant factors in the Senate's decision, the views of the directors of the three US nuclear weapons laboratories also contributed substantially to the outcome. Indeed, in their testimony of 7 October 1999, the directors offered a critique of the treaty that could well have led an uninformed listener to wonder why presidents since Dwight Eisenhower had sought such a treaty, and why the directors had agreed to President Bill Clinton's signing of it in 1996. As prominent voices on both sides of the controversy appear to have accepted the directors' assessment, an examination of their testimony is in order.

The directors all implied that to ratify the treaty now would amount to betting on unproven technologies: in the words of John Browne of the Los Alamos National Laboratory, "... we are using a fundamentally different set of tools to ensure the safety, reliability and performance of nuclear weapons", whereas "the essential tool kit for stockpile stewardship [without nuclear testing] will not be completed until the middle of the next decade". But this view is contradicted by the labs' own joint report¹ on how the stockpile had been maintained during the nuclear-testing era, which shows that the technology and methodology of that time were very similar to today's, and that underground tests played a minor role in establishing confidence in deployed weapons.

Paul Robinson of Sandia National Laboratory went further, recommending rejection of the treaty in a widely quoted statement: "If the United States scrupulously restricts itself to zero yield while other nations may conduct experiments up to the threshold of international detectability, we will be at an intolerable disadvantage. I would advise against accepting limitations that permit such asymmetry." This is a politico-military judgement in a technical guise, and is contradicted by testimony of the preceding day from General John Shalikashvili, who was chairman of the Joint Chiefs of Staff at the time of President Clinton's decision to sign the treaty. According to the general, the Chiefs had concluded — after consulting



Figure 1 Many thousands of parts in a warhead can be examined and tested without a nuclear explosion.

with the lab directors, among others — that on balance the CTBT was in the United States' interest. Moreover, the chief US negotiator of the treaty testified that, without the zero-yield limit, no ban would have been achievable, because the non-nuclear nations that the United States sought to constrain "insisted that the five [nuclear powers] should be allowed no tolerance — not even for the smallest possible nuclear yield"².

The question, of course, is what price the United States is willing to pay to constrain would-be nuclear nations. I will argue that, in view of the great strategic advantages that the United States enjoys over all potential proliferators, and indeed, all other nations, the benefits of the CTBT far outweigh the remote risk that abstaining from testing will generate a loss of confidence in the US nuclear arsenal.

Stockpile testing in the past

In assessing the confidence that the United States can have in its stockpile without nuclear testing, one should recognize that the great majority of nuclear tests were always devoted to developing new warheads, rather than reconfirming the reliability and safety of weapons that have been deployed for some time. From 1972 to the cessation of testing in September 1992, of the total of

about 350 underground nuclear tests³, only 68 were not devoted to development¹. Of these, 8 were "stockpile confidence tests" conducted on currently deployed weapons, and 35 others had direct application to these weapons (K. Johnson, Lawrence Livermore National Laboratory, personal communication).

Indeed, there is a logic to the paucity of tests following deployment: the human and technical delivery systems are much more likely to fail than the warhead itself; everything up to ignition of the fission reaction, including detonation of the conventional explosive and implosion, can be tested without a nuclear explosion (Fig. 1); the straightforward step of increasing the yield of the primary can be used to raise confidence in the performance of the secondary⁴ (see Box 1 on page 133 for an explanation of these and other terms); and the many thousands of parts in a warhead can be examined and tested in statistically significant numbers. In contrast, it would be prohibitively expensive to conduct enough nuclear tests to add a statistically meaningful measure of confidence to what was acquired from half a dozen development tests devoted to each new weapon design, followed by a test from an early production run to detect any error that might have crept in.

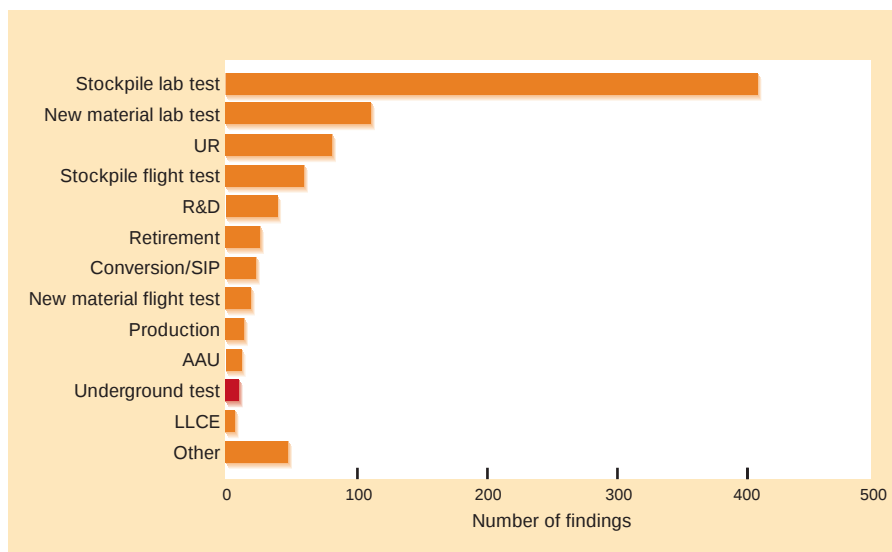


Figure 2 Activities in which “findings” of possible weapon defects were discovered from 1958 to the end of testing in 1992¹. About half of these findings led to modifications of the weapon type. UR, unsatisfactory report from the military services; SIP, Stockpile Improvement Program; AAU, Accelerated Aging Unit (thermal exposures to accelerate ageing); Underground test, underground nuclear explosive test for development, confirmation of production runs, stockpile confidence and to investigate nuclear effects on components; LLCE, Limited Life Component Exchange. “Stockpile” designates activities on weapons already deployed.

During the nuclear-testing era, stockpile maintenance already involved an extensive variety of destructive and non-destructive laboratory tests of components and sub-systems, as well as high-fidelity stockpile-to-target flight tests, in which the ‘physics package’ is replaced by a data-recording device that also simulates the weight and moment of inertia of the actual weapon¹.

The comparative utility of nuclear and non-nuclear tests is indicated by Fig. 2, which shows the various ways in which defects in nuclear warheads were found between 1958 and 1992¹. Of about 830 “findings” of potential defects and anomalies, only 1% came from all types of nuclear test. Note that this compilation includes findings in the new designs that were deployed after inadequate development testing, or even none, because of the testing moratorium from 1958 to 1961^{1,5}.

Statistics alone cannot, of course, rule out the possibility that the handful of findings from nuclear tests other than those related to the moratorium were of vital significance. But no knowledgeable person has made such a claim, and the record indicates that nuclear testing has played only a minor role in maintaining confidence in the modern stockpile.

Stewardship without testing

In signing the CTBT, the president in effect bound the United States to the provisions of the treaty, so only dramatic and unpredictable political or technological developments could lead the United States to resume nuclear testing, whether or not the treaty is ratified and comes into force. Furthermore, current US policy requires the weapons labs,

the Secretary of Defense and the military to certify annually all nuclear weapon types in the stockpile. This has been done since testing stopped, and makes it most improbable that new designs will be deployed in the future unless the de facto CTBT regime collapses.

As a result, the shelf-life of a warhead type can no longer be set by the tempo of replacement by more modern types, but will have to be indefinite. To meet this demanding requirement, resources and facilities are needed to refurbish or remanufacture any and all components of every warhead of a given type to their original and proven specifications when inevitable signs of deterioration appear, and a scientific staff having the requisite esoteric skills must be retained even though no new weapons are being developed and tested.

In facing this problem, the labs have placed their greatest emphasis on developing “a sufficiently detailed understanding of the science and technology that govern all aspects of nuclear weapons”, to quote Bruce Tarter, the director of the Lawrence Livermore National Laboratory. To this end, Congress is funding the science-based Stockpile Stewardship Program, a suite of diagnostic tools and computers that go far beyond what was available for developing any of the weapons in the enduring stockpile. A major purpose of these facilities is to provide the sophisticated tools that scientists of high calibre expect to have at their disposal.

Much of the work under the stewardship programme is devoted to phenomena that could affect the ageing of weapons, and to providing early warning thereof—a problem that needed much less attention when new designs were being introduced. But it is

hyperbole to claim, as did John Browne of Los Alamos, that “we are using a new method: a sequence of surveillance–evaluation–response. In this new paradigm we are using a fundamentally different set of tools.” Aside from the controversial and costly National Ignition Facility, which is of questionable relevance to stockpile maintenance, much of the Stockpile Stewardship Program consists of upgrades or sequels to earlier installations, and today’s methodology is not so different from that used for monitoring deployed weapons when testing was being done¹.

The weapons labs, and several other Department of Energy facilities, are responsible for weapons remanufacture. The directors have not, however, given remanufacture the emphasis that some expert observers believe it should have (see, for example, ref. 6).

The directors’ testimony

For decades, the position of the lab directors has been an important factor in both successful and failed efforts to constrain nuclear testing. Sandia’s Paul Robinson testified that this time the White House met the labs’ conditions after having asked, “what would it take to get you on-board [in support of the CTBT]?” These negotiations produced the Stockpile Stewardship Program, but apparently no enduring understanding. Robinson did not explain, nor did any senator ask, why he was now advising against ratification of the treaty.

It is natural and proper for the directors to keep the needs of their laboratories in mind when testifying. Weighing the technical issues in the light of the nation’s geopolitical interests is ultimately the responsibility of the president and the Senate. Nevertheless, widespread ignorance and anxiety about nuclear weapons imposes a responsibility on the directors not to stimulate misconceptions that are likely to flow from what they say.

Yet two widely held misconceptions arose directly from the directors’ testimony: that today’s stewardship programme replaces a programme that relied primarily on underground nuclear tests; and that this “new paradigm” could fail and put the US deterrent in doubt unless future computer simulations live up to promise on schedule. As described above, however, the essential features of the current stewardship programme were already in place when tests were still being conducted.

The directors largely ignored the enormous strategic transformation produced by the implosion of the Soviet Union, and that US nuclear weapons were designed to meet requirements set by Soviet targets. To varying degrees, they spoke as if the nation’s survival would be at stake unless there were an ever-ready ability to promptly begin testing of new weapon designs should some party “break out” of the test-ban regime, or should some dire threat suddenly arise which could

be countered only by a novel nuclear warhead that cannot be found either in the strategic arsenal or in the wide array of proven designs for tactical nuclear weapons. In fact, the CTBT imposes no constraints on means of delivery, so that a virtually unlimited range of new requirements for nuclear weapon systems can be met without reliance on new and untested nuclear designs.

The most implausible warning was Robinson's statement that "we will be at an intolerable disadvantage" should others conduct tests with very low yields while the United States adheres to the zero-yield limit. This issue was examined before the president signed the treaty, in a high-level technical study whose unclassified conclusions⁴ state that such low-yield tests would require modifications of warheads that would turn them into "at best a partial and possibly misleading performance indicator".

Of course, Robinson's fears raise more than a technical question. But the broader

Two widely held misconceptions arose directly from the directors' testimony.

issue was addressed by General Shalikashvili when he testified that the Joint Chiefs had concluded that "with a very mature and sophisticated nuclear program ... we would be better off to go to a zero-yield testing [limit], because it would have less of an impact on our stockpile" than on those of other countries "we're concerned about".

The price of rejection

All agree that nuclear proliferation poses a grave threat to the international order and to the vital interests of the United States. Although no one believes that the CTBT

is a magic bullet that will eliminate this danger, its benefits are very substantial and must be kept in focus. The existing verification system suffices to detect tests needed to develop sophisticated thermonuclear weapons (see Box 1), whether or not a state already has a fission weapon, or whether it seeks to turn stolen design information into a deployable weapon. Should the treaty come into force, it will create a much stronger verification system than the one now at work, which already sets a low ceiling on the yield of clandestine tests⁷. Only far-fetched scenarios can conjure up risks to the United States that are comparable to these benefits.

The United States formally committed itself to the CTBT in leading the campaign to gain an indefinite extension of the Non-Proliferation Treaty². The Senate's rejection of the CTBT has therefore done severe damage to the nation's ability to conduct diplomacy, for in essence the Senate has stated that the United States will not abide any shadow of doubt about its prodigious nuclear forces, while at the same time it continues to insist that the vital interests of non-nuclear states are best served by not acquiring such weapons!

The Senate's rejection makes it improbable that India and Pakistan will now sign the CTBT, and quite probable that these countries will resume testing to develop more reliable and destructive weapons. A nuclear arms race on the Indian sub-continent, given the proximity of China, would have global consequences. Nor can it be excluded that China, and perhaps even Russia, will resume testing, as both have more plausible reasons than does the United States to be concerned about abstention.

One must hope that the United States will be able to reverse course before rejection has such consequences. If that opportunity arises, the lab directors will have to conduct an embarrassing retreat from the branch on which they are now perched. ■

Kurt Gottfried is at the Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853-5001, USA.

Box 1: Nuclear weapons and nuclear testing

- The prospects for deploying a nuclear weapon without testing depend on the sophistication of the design, and on the proliferator's knowledge, resources and goals. Although the acceptance of risk is a subjective matter, two such different societies as the United States and the Soviet Union devoted comparable efforts to attaining a high level of confidence in their nuclear weapons. Here, confidence has two aspects: safety (assurance that a violent impact that detonates the chemical explosives, whether by accident or in warfare, will not cause a nuclear explosion or a dispersal of hazardous material) and reliability (knowledge that the weapon will perform to design specifications). Some conjecture that clandestine proliferators will not have such scruples, but, given their limited capabilities, they would have ample reason to ensure that their small arsenal is not a graver threat to themselves than to their enemies.

- The simplest fission weapon is of the Hiroshima type, in which a supercritical mass of isotopically purified uranium is suddenly assembled by a gun-like device. The Hiroshima design was not tested before its use, and presumably a

proliferator could repeat this feat, as did South Africa. But such devices have low yields relative to their weight and cannot use plutonium.

- All modern designs are descendants of the Nagasaki bomb, in which a sphere of fissile material is compressed by a surrounding shell of chemical explosives to form a supercritical mass. This device can use plutonium — either weapons-grade or from the spent fuel of a power reactor. It is a trickier design, however, and to bank on such weapons without nuclear testing would entail considerable risk of very low yield or outright failure.

- The next level of sophistication is the 'boosted' fission device, in which a mixture of hydrogen isotopes, placed at the very centre of an imploding fissile shell, undergoes a fusion reaction which produces energetic neutrons that greatly amplify the fission efficiency. The result is a far lighter weapon with yields that can reach hundreds of kilotonnes, but, without testing, it would be burdened by still greater doubts about reliability and performance.

- The five major nuclear-weapons states use much more complex, two-stage designs.

Here the energy released by a boosted fission device, called the 'primary', ignites a spatially separated thermonuclear 'secondary', giving a yield that can range up to many megatonnes. This whole 'physics package' involves more demanding computations and more sophisticated physics. Those building such a weapon for the first time, even with stolen information, would be taking great risks if they did not test.

- India and Pakistan advertised their tests in 1998, but they would have found it difficult to evade detection even though the monitoring system mandated by the CTBT was (and is) not in place. Although doubt has been cast on India's claim that it successfully tested a two-stage thermonuclear device, there is little question that it is working towards such a weapon, and that it will reach this objective if it continues to test^{8,9}. According to US experience, tests with a yield of 10 kilotonnes are required to establish confidence in such a weapon⁴. To conceal such tests would be a difficult undertaking for a proliferator, and attempts to do so could well fail, producing a detectable signature with untoward political consequences.

1. Johnson, K. *et al.* *Stockpile Surveillance: Past and Future* (Sandia National Laboratory (SAND 95-751/UCUC-700), January 1996).
2. Ledogar, S. J. *Testimony to Senate Foreign Relations Committee* (7 October 1999).
3. Mikhailov, V. N. (ed.) *Catalog of Worldwide Nuclear Testing* (Begell House, New York, 1999).
4. Drell, S. *et al.* *Nuclear Testing: Summary and Conclusions (unclassified)*. JASON Report JSR-95-320 (MITRE Corp., August 1995). <http://www.fas.org/rig/jsr-95-320.htm>
5. Kidder, R. E. *Maintaining the US Stockpile of Nuclear Weapons During a Low-Threshold or Comprehensive Test Ban* (Lawrence Livermore National Laboratory (UCRL-53820), October 1987).
6. Kidder, R. E. *Nature* **386**, 645–647 (1997).
7. Richards, P. G. & Kim, W.-E. *Nature* **389**, 781–782 (1997).
8. Wallace, T. C. *Seismol. Res. Lett.* **69**, No. 5, 386–393 (September/October 1998). <http://www.geo.arizona.edu/geophysics/faculty/wallace>
9. Perkovich, G. *India's Nuclear Bomb: The Impact on Global Proliferation* (University of California Press, 1999).

The truth is in here

How a science-fiction TV series treads the line of scientific plausibility.

The Real Science Behind The X-Files: Microbes, Meteorites and Mutants

by Anne Simon

Simon & Schuster: 1999. 320 pp. \$25

Henry Gee

There are two reasons why proponents of what is now called the Public Understanding of Science (henceforth PUS) find it so hard to get their signal to rise above the general noise of pseudoscientific psychobabble.

The first is a deficit of any sense of fun. I'm sure that people don't take astrology (to take one example) as seriously as PUS mavens think. Like dirty books were to Tom Lehrer, horoscopes are fun, and that's all there is to it — and, I dare say, generally more enjoyable than many scientific papers in these pages. Yet the puritans of PUS would have us avert our gaze.

The second, more serious reason why the PUS people have missed the mark is a failure to understand that science is a creative activity, driven by imagination. The generative phase of the scientific method — the hypothesis — rests on the imaginative leap enshrined in the question, 'what if...?'. What if I knocked out this gene rather than that one? What if I examined that remote galaxy at this wavelength?

It's no accident that an early diet of science fiction inspires many to embark on scientific careers. "Creative speculation seems to be the force that drives science forward, rather than the limiting conventions of accepted truth," says Chris Carter, creator of the hit science-fiction TV series *The X-Files*, in the foreword to *The Real Science Behind The X-Files* by Anne Simon — academic virologist, Carter family friend and science adviser to the series. So much is self-evident — if we were always content with the accepted truth, science would stagnate. Says Simon herself, "Critics that claim *The X-Files* is harmful to the public's awareness of science would probably be amazed to learn how many students in my freshman biology class point to the favorable portrayal of science and scientists on *The X-Files* as one reason for their interest in science".

The X-Files follows the adventures of a pair of FBI agents, Fox Mulder (played by David Duchovny) and Dana Scully (Gillian Anderson), as they investigate reports of paranormal or otherwise inexplicable activity. Mulder is a dreamer and a believer: if the existence of ghosts and ghouls is necessary to explain the facts of a case, then they exist. "The Truth Is Out There" is his motto. Scully,



Partners in the paranormal: but Mulder and Scully's investigations are grounded in science.

a qualified medical doctor, plays Watson to Mulder's Holmes — her task is to keep Mulder's feet on the ground and seek rational explanations for the various manifestations of eldritch and unspeakable horror that confront the pair each week.

The X-Files is tightly scripted and beautifully shot, but the best thing about it is that Mulder and Scully regularly fail to find any explanation for the phenomena that confront them — they must go back to base and come up with new hypotheses; the villain isn't always unmasked in the final scene. The frustration of forever chasing wild geese up blind alleys is much more like the real life of scientists than the neat, linear process of discovery of which PUS people would no doubt approve.

The action is seen through Scully's eyes. This, and the fact that she is portrayed as believable and normal (not to mention female), rather than an obsessive, antisocial geek, must be a minor milestone in TV, just as it was when *Star Trek* put a black woman (Nichelle Nichols, who played Uhura) on the racially integrated flight-deck of the starship *Enterprise* a generation ago. Indeed, like all good science fiction, *The X-Files* scores by placing its fantastic and often disturbing plots against a backdrop of detailed verisimilitude. The scientists have believable conversations, and use the right lab equipment. Says Carter, "I would tell the writers

and directors *ad nauseam* that the show was only as scary as it seemed believable, and only as believable as it seemed real or plausible". As adviser to the series, it is Simon's task to ground *The X-Files* in reality.

As such, *The Real Science Behind The X-Files* is an absorbing memoir for *X-Files* fans, illustrating, often in great detail, the scientific background to many of the series' more memorable episodes, and showing the degrees to which each episode veers from present reality. Simon writes in a bright, breezy and breakneck style, and manages to cram in an immense amount of detail about recent scientific research in a dazzling variety of areas — possibly much more than the average TV viewer would want, given that a little mystery at the edges adds to the general effect. It helps to have seen a good many episodes of the series before reading this book, and true *cognoscenti* will devour it — like SF fans in general, *X-Files* devotees dissect their favourite series with a microscopic assiduity that anyone else would find pathologically retentive.

This is, perhaps, a reason for sadness, for Simon writes about science in a refreshingly clear, direct and unpatronizing style that many readers not so enamoured of *The X-Files* would appreciate; particularly as she is unashamedly frank about where she thinks that the storylines stray from the

near-plausible to the inexcusably fantastic. The book as a whole should be an antidote to the general hostility with which *The X-Files* has been greeted by the PUS community, who condemn it as pandering to the public taste for pseudoscience. When Chris Carter was invited to address the Committee for Scientific Investigation into Claims of the Paranormal (CSICOP), he felt that his audience held him personally responsible for “all the loopy, looney trends in angels and aliens, in superstition, and even in fundamentalism”. It was as if he was “threatening to destroy yet another generation of minds by feeding them more bogus claptrap”. But Carter had his secret weapon in Anne Simon, who, like Carter’s accusers, “is a skeptic and, like Agent Scully, has a trust and faith in the scientific process”. CSICOP seems to have got the message — “I’ve never heard from them since,” says Carter. ■

Henry Gee is a senior editor at Nature.

Simplicity out of complexity

Strange Beauty: Murray Gell-Mann and the Revolution in Twentieth-Century Physics

by George Johnson

Knopf/Cape: 1999/2000. 432 pp. \$30/£17.99

Jonathan R. Ellis

Murray Gell-Mann is one of the towering figures in the field that uncovered a deeper and simpler layer of matter. George Johnson skilfully describes the fascinating train of events that led to the postulation and discovery of quarks, reconstructing the story around Gell-Mann’s central role, and counterpointing it with many insights into Gell-Mann’s complex personality.

Gell-Mann’s family background and childhood give us intriguing insights into his personality. His immigrant parents’ fortunes fluctuated; his father was apparently a hard taskmaster and his mother neurotic. Here we learn the true origin of the hyphen in his family name. Gell-Mann’s legendary polymath and polyglot precocity shone forth early. He skipped three years of school, and his remarkable high-school valedictorian address was featured in a cartoon in *The New Yorker*. He sped charmingly through undergraduate school at Yale and graduate school at MIT, bursting onto the particle-physics scene at the ripe old age of 21.

For most of the next two decades, Gell-Mann dominated the international world of high-energy physics. He played a leading role in the discovery of the strangeness quantum number, SU(3) symmetry and the eightfold way, quarks and current algebra — not to mention dispersion relations, V-A



Polymath and polyglot: Gell-Mann dominated the world of high-energy physics for 20 years.

theory, and much more. During this period, he worked very closely with experimental physicists, providing them with crucial guidance and helping to resolve the apparently conflicting results they sometimes obtained. Gell-Mann brought to all this both his light side of personal charm and his dark side of combativity. It culminated in his triumphal acceptance of the “Swedish prize”; his speech was made partly in the local lingo.

My own first encounters with Murray happened around this time, when I was a student — brilliant and inspiring lectures that cemented my ambition to attempt research in particle physics, the trickle-down of his deep insights into quarks and gluons via Bruno Renner, his collaborator and my advisor, and being charmed and flattered by the great man at our first personal meeting in a motel on the South Dixie Highway in Coral Gables. A year or two later, I made my post-doctoral pilgrimage to the Caltech ashram, but failed to hit it off with the guru. He was always surrounded by sufficient disciples, but by this time, field theorists were starting to have their revenge.

Gell-Mann’s attitude towards field theory had always been ambivalent, as he regarded it as a disposable means to an S-matrix end. He did not surf the gauge-theory wave as dominantly as the symmetric waves of yore, as Johnson explains in his account of Gell-Mann’s contribution to the birth of quantum chromodynamics. Here, by the way, the author takes one of his few scientific mis-steps: he wrongly identifies flavour SU(3) with colour SU(3). Also, he underestimates the extent to which the latter had been generally expected; and the importance of chiral symmetry in identifying

gluons as vector particles could have been underlined.

Gell-Mann continued to produce influential ideas for a while yet — for example, his co-discovery of the see-saw mechanism for light-neutrino masses in the context of grand unified theories, which is now so topical (although not mentioned by Johnson).

Murray’s subsequent forays into supergravity, superstrings and cosmology had less impact, but I am not competent to assess his later contributions to complexity theory.

Johnson tells us much about Gell-Mann’s foibles of character, including his penchant for learning languages and demonstrating this prowess, his encyclopaedic knowledge and inability to resist displaying it, and his sardonic wit. The complex relationship between Gell-Mann and his Caltech colleague Richard Feynman is less clear. It contained strong elements of rivalry, but also friendship and respect, shading into insecurity. A comparison of this book with *Genius* (Little, Brown, 1994), James Gleick’s biography of Feynman, inevitably comes to mind. *Strange Beauty* is certainly equally easy and interesting to read, and also quite thoroughly researched. *Genius* is perhaps stronger on psychological insight, and *Strange Beauty* may be stronger on scientific history. But these are details; anyone who enjoyed *Genius* will surely also appreciate *Strange Beauty*. The degree of appreciation may depend on the reader’s degree of empathy with Gell-Mann’s character.

Personally, I have fond memories of Murray — such as jostling to cede each other precedence on leaving a men’s room at Caltech (Murray won, of course) — and tremendous respect for his scientific achievements. As this book reveals, his personality is complex, and this may ultimately have restricted his scientific stature. He had a continual struggle with writer’s block, and his attention often wandered far from physics. If he had not put so much energy into knowing everything about everything, travelling everywhere, seeing every bird, consulting left and right, eating gourmet food, and other distractions, might he have achieved even more, getting up there with the greats of the millennium, not just of the century? His personal life might have been less interesting, but Gell-Mann might finally have been more satisfied. Setting aside these night thoughts, in this book you have a faithful record of one of the giants of late-twentieth-century physics — his life as well as his memorable science. This is a riveting read for anybody interested in the history and sociology of late-twentieth-century science. ■

Jonathan R. Ellis is in the Theoretical Physics Division, European Laboratory for Particle Physics, CERN, CH-1211 Geneva 23, Switzerland.

Of pigs, poultry and permafrost

Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus that Caused It

by Gina Kolata
Farrar, Straus & Giroux/Macmillan:
1999/2000. 256/330 pp. \$24/£12.99

David A. J. Tyrrell

Gina Kolata's book is not the usual sort of scientific review, even though it is based on wide reading of archival and published material and on many interviews with scientists in the field. Kolata studied microbiology, but has worked mainly as a journalist for *The New York Times*. As a result, the book, which outlines the science behind the influenza story, can be enjoyed both by general readers and by scientists.

Kolata describes the horrifying scenes of mass deaths that marked the 1918 influenza pandemic, and puzzles over the fact that it was referred to so little subsequently. Probably it was linked in memory with the horrors of the First World War that everyone wanted, and probably needed, to forget.

The epidemic in humans subsided, but it was noticed that outbreaks of a similar disease occurred that autumn in pigs in the American Middle West. Kolata includes agreeable pen portraits of important early figures, such as Richard Shope, who stumbled into growing the swine influenza virus, and she draws on the biography written of him by his friend Christopher Andrewes, who, with his colleagues Wilson Smith and Patrick Laidlaw, first grew a human

influenza virus in ferrets in 1933. Kolata gives a realistic description of the world of research — as a series of recurrent bafflements prompting various possible explanations, repeated probing and technical frustrations, then success that is quickly taken for granted — not at all the smooth course from initial question to complete solution that the general reader is fed by some newspapers and TV programmes.

Swine influenza comes into focus again as Kolata describes the 1976 outbreak of flu among young men in New Jersey. Scientific opinion at the time varied as to whether this was the beginning of another pandemic, although privately the chances were thought to be small — but not zero. So the question was, should the nation be vaccinated? This would require a formal report and politics at the highest level. Kolata infers that a consensus was manipulated by the politicians, and that alternative approaches and the possible adverse effects of vaccination were not considered in depth. The new vaccine was needed within months, and the government wanted the manufacturers to go ahead and also to obtain indemnity for possible harmful effects, but on this the insurance companies would not oblige. There was more delay. And then, as was forecast, when thousands of elderly people had been vaccinated, deaths began to occur. The deaths were probably coincidental, but

the newspapers started to produce "body counts" that were no doubt good for headlines but bad for public collaboration.

Although vaccination appeared to cause Guillain-Barré syndrome, Kolata suggests that the evidence for this was flawed and describes the problems of assessing complex scientific evidence in the law courts. The situation illustrates the difficulties of assessing risk when the

facts are being debated by the public, attention-grabbing journalists,

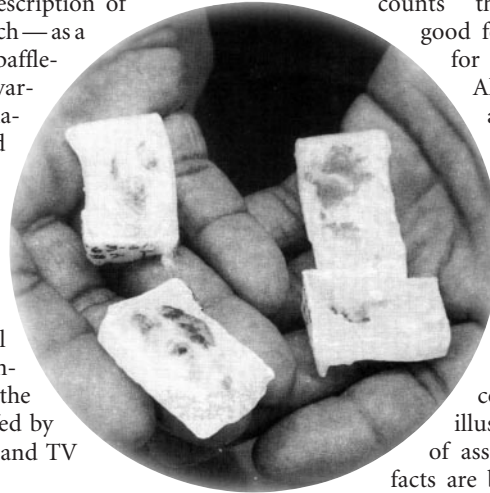
scientists, government and lawyers — Britain has had its share of this with BSE recently. It is a timely reminder of the need to develop an agreed 'framework of discourse', because something analogous to this will certainly arise again in the future.

Most enjoyable in the book is Kolata's description of the recent success of Jeffrey Taubenberger and his team in recovering and sequencing virus genomic material from fixed and embedded lung tissue taken from soldiers who had died in the 1918 pandemic. The tissue had been stored in the Walter Reed archive. The study originated in lateral thinking (the chapter is called "John Dalton's eyeballs"), since the team's main job was to develop improved molecular-based diagnostic methods for general use.

The story intertwines with that of another fascinating personal odyssey. In 1949, Johan Hultin was a young Swedish student who, with his wife, came to the United States. On his way to a studentship in Iowa, he visited Alaska and met a palaeontologist there. Later, he returned to Alaska to look for influenza virus in the bodies of victims of the 1918 pandemic buried in the permafrost. His expert local knowledge and good relations with the local people meant that he was able to visit suitable graves and obtain appropriate material. He failed, however, to recover the virus in the laboratory.

Hultin went on to become a successful pathologist and, in due course, retired. But he subsequently read about Taubenberger's new identification techniques, and within days was on his way back to Alaska, at his own expense. In spite of the further lapse of time since his last attempt, Hultin obtained lung tissue from the frozen bodies, from which Taubenberger identified the influenza virus genome.

Kolata also describes a large and expensive international collaboration, which was given heavy television coverage, to try to recover viral material from similar graves in Norway. It seems to have been unsuccessful



Deadly delivery: the 1918 pandemic struck down the young and fit. The top picture shows lung tissue samples, preserved in paraffin, from victims of the pandemic.

and it is clear which enterprise the author prefers.

Finally, Kolata brings us up to 1998, when a chicken virus of type H5N1 was recovered from human influenza patients in Hong Kong. The type had never been recovered from man before and thus was a potential pandemic strain. It turned out that there was an epidemic of the virus in the poultry markets in Hong Kong, fuelled by a constant influx of healthy birds from mainland China. The virus did not become established in humans and the outbreak was eliminated by slaughtering all the poultry in the territory. Possibly this action saved the world from another pandemic. But of course, successful public-health management, even on a world scale, gets little notice and less praise.

All in all, the book is a good read, but my advice would be to get a good textbook on influenza as well if you want the full picture. ■

David A. J. Tyrrell is at the Centre for Applied Microbiology and Research (CAMR), Porton Down, Salisbury SP4 0JG, UK.

Integrals of disease

Statistical Aspects of BSE and vCJD: Models for Epidemics

by Christl A. Donnelly & Neil M. Ferguson
Chapman & Hall: 1999. 229 pp. £39, \$69.95

Graham F. Medley

How should infectious diseases be managed? Typical answers to this question will include medical, social, economic and legal frameworks. However, quantitative study of what is essentially an ecological problem should also be expected. Any management plan to conserve a population should include a mathematical description of the expected changes in population size. Control of pests and parasites is on the other side of the same coin, so shouldn't a plan to control an infectious agent include a quantitative model of expected impacts and effects? The field begun by Roy Anderson and Robert May (*Infectious Diseases of Humans: Dynamics and Control*; Oxford University Press, 1992) is heading towards a more 'engineered' approach to the management of infectious diseases. Building a road without predicting its effect on traffic flow is irresponsible, so why implement a control programme without calculating its expected effects?

The UK government was culpable for the lack of appropriate, published quantitative study in its handling of the BSE epidemic. Anderson and colleagues published the first detailed understanding of the infection process and intervention consequences in 1996 (see *Nature* **382**, 779–788; 1996). Until then, the impression, if not the reality, was of

Science in culture

The life of Paul Dirac

The Oracle of Delphi, a show performed at CERN, Geneva

Alison Abbott

Deep under the European Laboratory for Particle Physics — CERN — in the huge space provided by the cavernous loading bay that serves LEP (the Large Electron–Positron collider) and its associated DELPHI antimatter experiment, the life of Paul Dirac is being played out nightly in an impressionistic show of acrobatics, dancing and light effects.

Dirac, the Nobel prizewinning theoretical physicist who predicted the existence of antimatter in the 1930s, writes to Werner Heisenberg about his struggle to understand the nature of matter. His muse helps him untie the knots in his mind and derive his mathematical proof for antimatter. His dreams take him into space to search for the antimatter he believes must be there, as he finds no proof for it on Earth. His disappointment and self-doubt — “if only I could test antimatter close-by” — turn to wonder when, in the last of four scenes, the backdrop suddenly opens to reveal DELPHI, the experiment that allows experimental physicists to study the interactions of electrons and their antimatter counterparts, positrons.

The Oracle of Delphi was created as a CERN outreach activity and has been enormously successful. It is almost sold out to the end of its run, mostly with locals from Geneva. Before the performance, the audience is given a short lecture in particle physics, and then descends to the loading bay — 100 metres below the ground — for the show. Afterwards, they are taken around a LEP experiment. The show's artistic success, which is certainly helped by the drama of



the space in which it is performed, is in part due to a concise script by Anne Gaud McKee, a molecular biologist from the University of Geneva, and the professional acrobats and mime artists who perform. ■

The Oracle of Delphi will run until 27 February 2000 (when LEP, currently switched off for winter maintenance, comes back into operation).

Alison Abbott is senior European correspondent for *Nature*.

quantitative management guided by intuition, guesswork and, I suspect, hope.

This book adds to the quantitative methodology that should underpin the control of diseases that have a long and variable incubation period. These diseases can be especially problematic because of the need for accurate estimates of the distribution of the incubation period in order to infer the temporal pattern of infection from disease data (back-calculation) or to predict the future epidemic from a known infection dynamic. That we have no precise knowledge of either the infection process or the incubation-period distribution of new variant CJD is the principal reason why accurate predictions of the CJD epidemic will be impossible until the epidemic is almost over and the incubation-period distribution has been estimated.

Integrals, with dimensions of age, time and time since infection, abound in the book, and one of the most remarkable features is the lack of statistics. A slightly more appropriate title might have been *Integral*

equations for diseases with long incubation periods: application to BSE and vCJD, but this would have been even less attractive than the given title. This slant is not surprising because, as the authors point out, the melding of statistical methods and nonlinear modelling requires considerable development generally, and the authors' rigorous contribution is to be welcomed.

The book provides a comprehensive view of BSE modelling as seen from the backroom. It pulls together the authors' published work, on back-calculation, maternal transmission and spatio-temporal clustering. However, there is little generalization beyond BSE and new variant CJD. And, disappointingly, the decision-making dimension is absent. Some discussion of the impact of this work would have broadened the book's potential audience to include those working in sectors that should actually be using mathematical modelling. ■

Graham F. Medley is in the Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.

Positioning the goalposts

The best environmental policy depends on how you frame the question.

John Maddox

Is sustainability sustainable? We are all environmentalists now, of course, and since the Brundtland report in 1992 we have hewn to the doctrine that the only good, or even permissible, changes in the Earth's surface and the uses we make of it are 'sustainable'. Exactly what sustainable development means is, unfortunately, not crystal clear. On the strictest reading, the burning of a single tonne of fossil fuel would be banned, because the result would be that coal and petroleum reserves were less easily accessible to future generations, leading to the conclusion that the only permissible energy sources are renewable — Sun, wind and water.

Not even Greenpeace or Friends of the Earth go quite as far as to advocate an immediate shut-down of the fossil-fuel industries. It is not practical politics. In any case, this strict reading of sustainability leads directly to a paradox: it would deny to future as well as present generations the exploitation of any finite resource, meaning that future generations would be no worse off if the resource had already been consumed by the time they came on the scene. One way round that difficulty is to relax the reading of sustainability.

Petroleum, for example, now sells at \$22 a barrel, so we can claim that we are exploiting only the reserves whose production cost is less than that. So far as we know (for the oil companies have no incentive to look for uneconomic reserves), there will be plenty of \$30-a-barrel oil left in the ground a decade hence. We cannot now know whether future generations will be able to afford steadily rising oil prices, of course. The market will in any case decree that the cheapest source of energy be exploited first.

Either the next generation will pay \$30 a barrel for oil, or it will use renewables or nuclear energy. Either way, loosening the definition of sustainability entails the exhaustion of current resources and their replacement by others. How long can that continue? The missing ingredient in all arguments about sustainability is that there is no agreement about long-term goals — or even on what is meant by 'long' term. The two questions are inseparable and inevitably in conflict, and for a simple reason.

To a large degree, *Homo sapiens* has opted out of the tyranny of natural selection, often for the best compassionate reasons. Since the beginning of civilization, we have protected ourselves against the elements by building shelters and have fed ourselves by growing



Here today: there is no general agreement about the meaning or timescale of sustainable development.

crops and herding domesticated animals, transforming the ecology of large tracts of the Earth's surface in the process. Now we offer infants paediatric medicine and do what can be done to avoid the deaths of adults, even those beyond reproductive age. One consequence is that the human population (now in excess of 6 billion people) is much greater than that of comparable animals at the tops of food chains — whales or tigers, for example. Another is that welfare rolls are growing in many countries, rich countries in particular. A third is pollution on a global scale, such as the accumulation of atmospheric carbon dioxide.

Responding to such problems requires prior decisions on objectives. Should global policy aim at the indefinite preservation of the human race, as seems to have been implicit in human behaviour since the first human settlements 5,000 years ago? And, if so, does that mean the survival of 6 billion people or some other number? And do we aim to hold to the goal for another century, for a millennium, until the Sun becomes a red giant or even for the rest of time?

It would be simplest if none of these questions required an immediate answer. Even if, as seems probable, the Intergovernmental Panel on Climate Change has exaggerated the

impending rate of temperature increase by a factor of two, our misfortune is that some kind of determination will probably be necessary in the next century or so. If the objective is the survival of the human race, we would surely be spending more on carbon dioxide abatement, less on biodiversity and a lot more than we do at present on avoiding asteroid impacts with the Earth. If survival beyond the red-giant phase of the Sun is taken seriously, Mars would be colonized sooner than would otherwise be prudent.

The pity is that there is no forum in which these matters can be discussed, let alone decided. But this is probably just as well. The admirable United Nations is good at making compromises, but they are usually expensive — so expensive that they drain funds from other urgent tasks. And the obvious determinations of the key questions of whether the survival of the human race should be our goal, and whether the timescale is indefinitely long, provoke illiberal solutions — eugenics for all at the very least.

So we must fall back on the old-fashioned remedy of public argument and controversy. It is to be hoped that the next millennium, even the next century, will see the need. ■

Sir John Maddox is Emeritus Editor of *Nature* and is at 9 Pitt Street, London W8 4NX, UK.

Deep Ice and DNA languages

Life, but not intelligence, abounds under the ice of frozen planets.

Greg Bear

2215 CE — The discovery of more than 15,000 Massive Deep-Ice Objects (MDIOs) in orbit around supermassive planets in the close interstellar neighbourhood has the potential to revolutionize our understanding of biological languages.

Phenotype-generating languages for terrestrial DNA-recorded life forms can be ranked closely according to kingdoms. Most plants, for example, express phylogeny according to the famous Zinn–Wang languages, first decoded in the mid-twenty-first century.

Archaea, commonly used as the Rosetta Stone for all primitive DNA languages, have provided deep insights into nonterrestrial biologies that have advanced to the DNA level.

RNA languages in terrestrial viruses constitute a virtual Tower of Babel, indicating a degenerate and mutationally rich mix that can still compel replication in DNA hosts. Early RNA coding systems found outside the Solar System, however, can often be translated into Archaean DNA-based languages, and these may constitute the most basic fixed languages of life.

MDIOs are typically seven to nine Earth masses and consist of layers of water ice, 2,000–5,000 km in depth, overlying a high-pressure liquid ocean (HPLO) that sits in turn on a rocky mantle. At the centre is an iron- and sulphur-rich core.

The interior is warmed both by latent heat and radioactivity (chiefly thorium-90) and by tidal friction generated by interaction with the parent supermassive body. Temperatures in the HPLO can frequently exceed 130 °C. MDIOs may constitute the most common life-supporting bodies in the Universe.

Neutronium self-guided probes (NSGP) have penetrated seven of these ‘deep ice’ objects and have obtained remarkable data. Spin-off probes made of normal matter, transmuted from neutronium and released into the HPLO, perform *in situ* analysis of all carbon chemistry and send information to orbiting research stations.

What is most remarkable about MDIO biologies is that they can exist at all under these extraordinary conditions. Once again, it is demonstrated that life will begin and thrive anywhere there is liquid water and the necessary elements.

To date, RNA- and DNA-language analysis has been conducted on three HPLO ecosystems. Because of limited exploration



ranges for the probes, the extent of these ecosystems is not known: however, every probed MDIO possesses life.

One of the ecosystems (MDIO 2341a) is still in a ‘profligate’, mutation-rich RNA phase, with no complex organisms and no DNA detected. Here, new genetic coding schemes may naturally emerge every few months, and competition between coding schemes is likely to be extreme. (The emergence of competent and stable genetic languages on Earth may have taken more than a billion years.)

The other two ecosystems (MDIO 5756b and MDIO 349x) have entered the more stable age of DNA and show remarkable similarities to each other.

The most striking feature of these ecosystems is how bright they are, since they are completely hidden from all starlight. The roving probes have sent back images of massive reef-like structures glowing as brightly as several full moons, surrounded by a thick, slowly churning mass of light-dependent microbes. Feeding on these microbes are living filter nets, fringed by corkscrew cilia, able to join into larger units or separate into smaller ones.

Tall spiral chimneys, like baroque columns in a church, release water heated in the upper mantle, creating plumes that can extend for many kilometres. These plumes spread out at the upper ice layer, eroding smooth domes almost a hundred kilometres in diameter and usually less than a millimetre deep. These domes collect oxygen freed from the actions of photosynthetic organisms. Typically the oxygen is forced back under extreme pressure into the water and the ice within seconds, but during this brief time, miniature forests of opportunis-

tic organisms grow in the domes, extracting all the benefits from a more efficient oxygen metabolism.

The upper limit of organization in MDIO ecosystems is not known. Typical distributed-intelligence ecosystems are found here, and interact to form complex neuronal networks that govern MDIO life cycles (as on Earth). However, no condensed nodal intelligences such as large animals have yet been found. Instead, intelligence seems frozen at a very early and distributed stage.

This may reflect the unlikelihood of any intelligences within the MDIO ever being challenged by major environmental change, much less being given a chance to observe the outside Universe. (MDIOs seem to be remarkably stable over hundreds of millions of years.)

The impossibility of emergence through the deep ice and escape into space limits the potential growth of concentrated hypothesis engines as defined by the Turing–Watteau diagram of novel information versus expansion opportunity.

Some researchers suggest that the seeding of provocative artefacts (‘Clarkeing’) below the deep ice may encourage condensation of concentrated intelligences, or, at the very least, induce some interesting emergent properties. The design of these artefacts is currently spurring intense debate. As one chief communications researcher has asked, “How do you uplift slime?”

Harnessing of bacterial communities on Earth in the last century could provide a template for working with MDIO ecosystems, adding to the list of beings we can actually talk to.

Greg Bear is the author of 26 novels, the most recent being *Darwin's Radio* (HarperCollins).

DNA computing on a chip

Mitsunori Ogihara and Animesh Ray

In a DNA computer, the input and output are both strands of DNA. A computer in which the strands are attached to the surface of a chip can now solve difficult problems quite quickly.

“How often have I said that when you eliminate the impossible, whatever remains, *however improbable*, must be the truth?” exhorted the great sleuth¹. The principle of arriving at the truth by elimination is ancient; but on page 175 of this issue, Liu *et al.*² report a new technique for massively parallel elimination, which harnesses the power of DNA chemistry and biotechnology to solve a particularly difficult problem in mathematical logic.

The difficulty of finding solutions to mathematical problems is classified by the speed at which the best algorithm can compute their solutions. ‘Easy’ problems have algorithms with ‘running times’ that scale as a polynomial function of the number of variables (polynomial time or P problems). There is also a class of problems characterized by proofs that are easy to verify (non-deterministic polynomial time or NP problems), such as the famous travelling salesman problem. In the worst case, ‘hard’ NP problems have running times that grow exponentially with the number of the variables. For example, finding a factor of a given natural number N cannot be done in polynomial time, but verifying that another number d is a factor of N is easy. Computer scientists have been intensively studying whether sequential algorithms can solve all NP problems in polynomial time, but the answer is still unknown.

In 1994, Leonard Adleman³ shocked the computing world by presenting a DNA-based polynomial-time method for the Hamilton path problem (Fig. 1a), the problem of finding an airline flight path between several cities on a map such that each city is visited only once. This NP problem is known to be one of the hardest. In order to achieve the small computation time, Adleman traded space (the amount of DNA needed) for time (the number of biochemical steps to be used). His key insight was that cities on a map, and paths between pairs of cities, may be encoded in strands of DNA. Millions of DNA strands, diffusing in a liquid, can self-assemble into all possible flight-path configurations, from which a judicious series of molecular manoeuvres can fish out the correct solution. Adleman, combining elegance with brute force, could isolate the one true solution out of many possibilities.

Every NP problem can be seen as the search for a solution that simultaneously

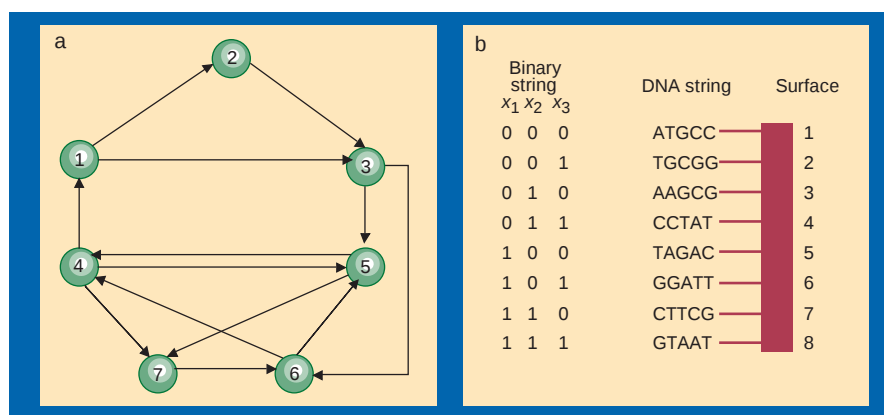


Figure 1 The parallel power of DNA computing. a, An example of the Hamilton path problem solved by Adleman³. Can you go from node 1 to node 7 using only the paths shown such that you visit all the nodes exactly once? The answer is positive. b, The hardest of such computationally difficult or NP problems is 3-SAT. In order to find a solution to the 3-SAT problem defined by these two clauses $(x_1 \text{ OR } x_2 \text{ OR } \bar{x}_3)$ AND $(\bar{x}_1 \text{ OR } x_2 \text{ OR } \bar{x}_3)$, Liu *et al.*² attach DNA strings encoding all possible answers to a specially treated surface. Complementary DNA strands that satisfy the first clause are added to the solution, and stick to strands numbered 1 and 3–8. The remaining single strand 2 is destroyed by enzymes. The complementary strands are removed and the surface is washed. The cycle is repeated for the second clause, which results in the destruction of strand 6. The identities of the remaining strands are read out to give the correct solutions to the problem: 000, 010, 011, 100, 110, 111.

satisfies a number of logical clauses, each composed of three variables (which can be true or false), connected by ‘or’ statements: for example $(x_1 \text{ OR } x_2 \text{ OR } \bar{x}_3)$ AND $(\bar{x}_1 \text{ OR } x_2 \text{ OR } \bar{x}_3)$. This particular problem, known as 3-SAT, is the hardest of all NP problems. Liu *et al.*² show how to solve a simple case of 3-SAT in a reasonable amount of time by using a brute-force search made possible by the parallel nature of their DNA computing techniques.

They begin with a string of binary numbers representing the variables in a given 3-SAT formula. Such a binary string can be represented by a unique sequence of nucleotides in single-stranded DNA; for example, TGCGG might stand for 001. For n variables, there are 2^n unique answer (or Watson) strands, so for three variables you need eight Watson strands. For each Watson strand, there is also a complementary Crick strand created by the base-pairing rule — A bonds to T, and C bonds to G. The goal is to identify those strings out of a library of eight that satisfy all the clauses of a particular 3-SAT formula (Fig. 1b).

Liu *et al.*² first immobilized the Watson DNA strings corresponding to all candidate solutions on a specially treated gold surface.

Next they added all possible Crick strands that will stick to a Watson string satisfying the first clause. Such pairing creates double-stranded DNA. The remaining single-stranded molecules are those that do not satisfy the first clause, and these are destroyed by enzymes. The surface is then heated to melt away the complementary strands, washed and a fresh collection of Crick strands is paired with strings satisfying the second clause. This cycle is repeated for each of the remaining clauses. At the end, only those strands whose sequence satisfies the original formula survive.

In this system, the DNA ‘answers’ are attached randomly to the surface (rather than in an ordered array) so, to read out the answer, the surviving strands first have to be amplified using the polymerase chain reaction. Their identities are then determined by pairing with an ordered array of strings identical to the original set of sequences. Not counting the number of steps required to produce the DNA molecules in the first place, the algorithm takes only $(3k + 1)$ steps, where k is the number of clauses, for a brute-force evaluation of all 2^n possible answers. This represents a remarkable improvement over the best conventional

computer algorithm⁴, which scales as 1.33^n , where n is the number of variables. For example, a 3-SAT problem with 30 clauses and 50 variables could be solved in approximately 1.6 million steps by an ordinary computer algorithm, but in only 91 steps by Liu and colleagues' DNA computer.

From a computer-science perspective these results are remarkable, but scaling up this technique to solve larger 3-SAT problems is still unrealistic. The authors claim that their surface-based approach is designed to scale up to problems of practical significance, but solving a problem that would embarrass a conventional computer is a long way off. First, there is the important issue of correcting errors arising from the inherent sloppiness of DNA chemistry, although ways of dealing with this issue have been suggested⁵. Second, there is the high cost of tailor-made DNA sequences. For example, a 50-variable 3-SAT problem will require at least 10^{15} unique DNA strands. It remains to be seen whether massively parallel synthesis of DNA binary strings, using combinatorial chemistry, could further speed up the technique and reduce costs. One also needs to consider the non-trivial problem of designing enough unique DNA strands that do not interfere with each other⁶.

The most serious remaining issue is the exponentially increasing number of DNA molecules needed to compute even small 3-SAT problems⁷. Alternative strategies that use self-assembly of complex DNA nanostructures also suffer from similar weaknesses⁸. Perhaps a compromise may be achieved by reducing the search space through heuristics⁹. An alternative may be to examine whether a surface-based attack on the more

manageable P problems, such as Boolean circuit evaluation, would yield better returns in terms of speed and cost¹⁰. If the DNA strings are spatially encoded and attached to an organic semiconductor surface, computation and read-out may also be combined.

It is foolish to attempt to predict the future of technology, but it may be that the ideal application for DNA computation does not lie in computing large NP problems. Some day there may be a need for fully organic computing devices implanted within a living body that can integrate signals from several sources and compute a response in terms of an organic molecular-delivery device for a drug or signal. Progress in DNA computing may pave the way in this direction.

Mitsunori Ogiwara is in the Department of Computer Science, University of Rochester, Rochester, New York 14627-0226, USA.
e-mail: ogiwara@cs.rochester.edu
Animesh Ray is at Akkadix Corporation, 11099 North Torrey Pines Road, Suite 200, La Jolla, California 92037, USA.
e-mail: ray@akkadix.com

1. Conan Doyle, A. *The Sign of Four in The Complete Sherlock Holmes* (Barnes & Noble, Dayton, New Jersey, 1988).
2. Liu, Q. et al. *Nature* **403**, 175–179 (2000).
3. Adleman, L. M. *Science* **266**, 1021–1024 (1994).
4. Schöningh, U. in *Proc. 40th Ann. IEEE Conf. Found. Comp. Sci. (FOCS)* 410–414 (IEEE Comp. Sci., Los Alamitos, California, 1999).
5. Adleman, L. in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* Vol. 44 (eds Lipton, R. & Baum, E.) 1–21 (Am. Math. Soc., Providence, Rhode Island, 1996).
6. Frutos, A. G. et al. *Nucleic Acids Res.* **25**, 4748–4757 (1997).
7. Adleman, L. M. *Sci. Am.* **279**, 54–61 (1998).
8. Winfree, E., Liu, F. R., Wenzler, L. A. & Seeman, N. C. *Nature* **394**, 539–544 (1998).
9. Ogiwara, M. & Ray, A. in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* Vol. 48 (eds Rubin, H. & Wood, D.) 255–264 (Am. Math. Soc., Providence, Rhode Island, 1999).
10. Ogiwara, M. & Ray, A. *Algorithmica* **25**, 239–250 (1999).

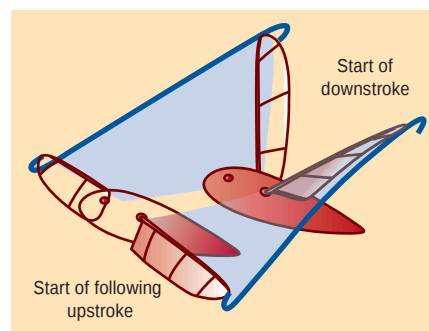


Figure 1 A possible micro-air-vehicle, following the principles of insect aerodynamics outlined by Ellington¹. Shown here are the start of the downstroke and of the following upstroke. As the stroke reverses the sail-like wings twist, so creating useful aerodynamic force at all stages of the wing beat. The area swept out by the wings, shaded blue, is here concentrated above the body, generating a forward force that balances the vehicle's tendency to tip backwards owing to the position of the centre of mass. Changing the position of this swept area would change the body angle and with it the flight speed.

component to generate forward, backward or sideways thrust as needed. To achieve this, insects oscillate and twist their wings, and typically vary the direction of the airflow by altering the angle between the plane of the wing stroke and the horizontal: the greater the angle, the faster the flight. The wings themselves deform semi-automatically, optimizing aerodynamic forces during the cycle².

Flight stability is maintained, and manoeuvres initiated, by finely controlling the details of the stroke. An asymmetric change in amplitude, or in the timing of twisting of the wings on one side of the body, can initiate a turn. An upward or downward shift of the line of action of the mean aerodynamic force can tilt the insect forward or backward, so altering the stroke-plane and in turn the speed.

Mimicking all these features in tiny vehicles presents many challenges: light enough materials; a miniature power plant; sophisticated controls; a finely controllable mechanism to flap and twist the wings at appropriate frequencies; and wings of appropriate shape, size and flexibility to maximize lift for minimum energy expenditure.

Two decades of research, much of it by Ellington and his group, have shown how insects make wide use of unconventional aerodynamic mechanisms which exploit the unsteady airflow resulting from the flapping motion to obtain far more lift than conventional, steady-state aerodynamics would allow^{3–5}. Some at least would be applicable to MAVs, particularly the 'dynamic stall' phenomenon⁴: as the wing accelerates into the stroke, an intense vortex develops above the leading edge, and spirals outward towards the tip. This allows the wing to

Aerodynamics

From insects to microvehicles

Robin Wootton

Insects are the world's smallest and in many respects most perfect flying machines. Many can hover, fly slowly and manoeuvre with great precision. They can navigate through narrow gaps and around complex spaces, constantly receiving detailed information about their surroundings — but keeping it to themselves. How useful it might be, in surveillance, industrial fault location and so on, if we could build little machines with similar capabilities to detect and transmit information to us.

Advances in our understanding of insect aerodynamics, energetics and structural mechanics have brought this objective within reach. Collaborative research between engineers and specialists in insect flight is under

way in at least three countries to develop micro-air-vehicles (MAVs) capable of this kind of investigatory flight in enclosed situations. Charles Ellington, of the Zoology Department, University of Cambridge, has published an intriguing account of the aspects of insect flight most applicable to MAV development, predicting how such a machine might best be designed¹.

In technology, the rotary aerofoils of helicopters are the traditional solution to slow, manoeuvrable flight and stationary hovering, but at the insect/MAV scale flapping is a viable alternative. Both systems operate by accelerating a mass of air, predominantly downwards to support the weight, but with a smaller backward, forward or lateral

operate at high angles of attack at which it would otherwise stall, and to gain substantial extra lift to support the insect's weight.

How should a MAV be designed? Most insects have two pairs of wings, often linked into a single aerofoil. Flies get by with one pair and so, with care, could the MAV. The wings must be deformable, at the least developing a propeller-like twist, which reverses between the half strokes, like the sail on a tacking dinghy. A simple sail-like wing with a controllable boom and a membrane might be a reasonable initial design. As a minimum, Ellington suggests that the stroke amplitude should be variable and controllable. So should the location, relative to the body, of the area through which the wing sweeps (Fig. 1), because this would influence the line of action of the mean aerodynamic force, which in turn would control the body angle, the stroke plane angle and hence the flight speed. The wings' angle of attack must simultaneously alter, to remain optimal as the speed changes. For manoeuvres, the amplitude of the stroke and the location of the swept area must be independently controllable on the two sides of the body.

Flapping is potentially expensive, as the wings must gain and lose kinetic energy during each half stroke. Insects minimize the cost by storing the energy elastically at the end of each half stroke and returning it at the start of the next. Micro-air-vehicles would need to do the same, operating as resonators oscillating at their natural frequency, which would necessarily be fairly constant.

In any flying machine the ratio of power to mass is crucial. It would be hard to make a machine as light as an insect of the same size. A tiny internal combustion engine or a lithium battery might well provide enough power input. How could power output be maximized? The principal available vari-

ables are stroke amplitude and frequency, and wing length. Amplitude should be left adjustable as a means of varying power for manoeuvres and perhaps for carrying loads. Wing length and stroke frequency would be built into the machine's design. Which would have the greater influence on power, and what else may they affect?

Ellington's analysis considers a hovering machine with a basic stroke amplitude of 120° , a lift coefficient of 2, and wings with an aspect ratio (span/mean width) of 7; these values are all well within the insect range. Flapping is assumed to follow simple harmonic motion, and the geometric centre of the wing to be halfway along its length. Using his own aerodynamic equations⁶, with a new assumption derived from the spiral vortex, he concludes that longer wings will be far more effective than higher frequency in raising the power/mass ratio, so that a given mass will be supported with less power by using longer wings and a lower frequency. On the other hand, higher frequencies would give higher maximum speeds.

The insects show what can be achieved given a good power source, clever controls, superlative materials and 350 million years of research and development. Ellington's analysis suggests how this might best be copied in a rather shorter time. Over now to the engineers. ■

Robin Wootton is in the School of Biological Sciences, University of Exeter, Hatherly Laboratories, Prince of Wales Road, Exeter EX4 4PS, UK.

e-mail: R.J.Wootton@exeter.ac.uk

1. Ellington, C. P. *J. Exp. Biol.* **202**, 3439–3448 (1999).
2. Wootton, R. J. *J. Exp. Biol.* **202**, 3333–3345 (1999).
3. Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* **305**, 79–113 (1984).
4. Ellington, C. P., van den Berg, C., Willmott, A. P. & Thomas, A. L. R. *Nature* **384**, 626–630 (1996).
5. Dickinson, M. H., Lehmann, F.-O. & Sane, S. P. *Science* **284**, 1954–1960 (1999).
6. Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* **305**, 145–181 (1984).

Astronomy

Eyes wide shut

David Jewitt

Astronomers like to forget that the roots of their subject lie in ancient superstitions about the influence of the cosmos on everyday affairs. In fact, astronomy and astrology were closely intertwined as recently as four centuries ago, when Tycho Brahe laid the foundations of modern astronomy while simultaneously maintaining a lucrative business in personal horoscopes. Modern astronomers generally scoff at such superstitious beliefs, so it is somewhat ironic that science has in the past few decades uncovered compelling evidence for celestial interference in terrestrial matters.

It is now clear that asteroids occasionally

wander from the main belt beyond Mars because of chaotic instabilities caused by Jupiter. Some of these errant asteroids strike the Earth with terrible consequences. On page 165 of this issue, Rabinowitz *et al.*¹ report that the number of threatening near-Earth objects (NEOs) larger than 1 km in diameter is only half the previous estimates. But we still have no effective means of detecting them all, and no form of self-defence.

The Earth bears the scars of previous encounters with NEOs. Hundreds of impact craters, some the size of small American states, have been discovered on the surface of our planet. Each was produced by a



100 YEARS AGO

For several years Prof. W. O. Atwater has been engaged in investigations to determine whether the energy given off from the body of a man in the form of heat, or of heat and external muscular work, is equal to the potential energy or heat of combustion of the material actually burned in the body; in other words, whether the law of conservation of energy holds good for the living organism. The latest number of the *Physical Review* (vol. ix., No. 4) contains a concluding account, by Prof. Atwater and Mr. E. B. Rosa, of experiments made with the view of testing this point. ... The mechanical efficiency of a man was determined by a comparison of the energy used when at rest and when performing muscular work. The work done, divided by the total energy yielded by the body, gave 7 per cent. as the mechanical efficiency. As, however, a large amount of the energy received was used up in the body, only the excess of energy absorbed in the work experiment over that required when the subject was at rest should be charged against the work done. When this was taken into account the mechanical efficiency of man came out at 21 per cent., which equals or exceeds that of the best compound condensing engines with the highest efficiency boilers. From *Nature* 11 January 1900.

50 YEARS AGO

As normal blinking is involuntary, it might be expected that blinking would proceed at its normal or at a somewhat reduced rate during sleep, as with breathing and heart-beat. This is not so, however, though little seems to have been published on the subject, especially on the quantitative side. ... In all cases so far examined, it has been found that in sleep the eyelids are quiescent, and show no signs whatsoever of blinking movements. During a recent long train journey, opportunity was afforded of examining the problem in its quantitative aspects, the subjects being a young man and his wife, both of whom fell asleep several times during the course of the journey. ... The interblink periods of each subject when awake and when in the resting condition were sensibly identical, whereas in sleep bilateral lid movements ceased entirely.

From *Nature* 14 January 1950.

devastating explosion that must have been fatal to life in the surrounding areas on scales from local to global (Fig. 1). The Cretaceous–Tertiary mass extinction of 65 million years ago seems to have been triggered by the impact of an asteroid 10 km in diameter². Ten thousand people killed by ‘falling stones’ in Shanxi Province, China, in 1490 were possibly the victims of a much smaller and thoroughly fragmented projectile. Still more recently, on 30 June 1908, 1,000 square kilometres of Siberian pine forest in Tunguska were blown flat by a 10-megaton atmospheric blast caused by a 70-metre asteroid.

The gradual acceptance of the evidence for impacts by asteroids (and comets) has led naturally to questions about the magnitude of the threat posed by NEOs to life on Earth^{3,4}. Rabinowitz and colleagues¹ provide the most recent and best controlled estimate of the number of large, potentially Earth-threatening NEOs. They report that there are nearly 1,000 NEOs larger than 1 km in diameter and that, given the present rate of discovery, it will take 20 years for 90% of these objects to be found. Should we worry?

The answer depends on the number of fatalities to be expected, but also on personal assessments of risk. The number of NEOs found by Rabinowitz *et al.* is within a factor of two of previous estimates based on less-controlled samples, so published estimates of impact mortality are essentially unchanged. Considering events of all energies there is about 1 chance in 20,000 of being killed by an impact during the course of a human lifetime⁴, similar to the likelihood of being killed in an airplane accident. The perception of risk from impacts is smaller than for being killed in a plane crash because planes crash at a steady rate with (relatively) few deaths per event, whereas lethal impacts are rare but kill a lot of people. At the very least, the potential consequences of impact are large enough to cause concern.

In the past decade, thanks to several reported near-miss encounters with small objects, the impact threat has become a subject of intense interest to the general public (spawning the popular movies *Deep Impact* and *Armageddon*). In 1994, the United States House Committee on Science and Technology went so far as to order the US space agency NASA to “catalogue within 10 years the orbital characteristics of all (Earth-orbit-crossing) comets and asteroids that are greater than 1 km in diameter”. This particular cut-off diameter was picked in part because 1-km NEOs are thought to be the smallest objects capable of wreaking global havoc (for example, by disrupting the climate and shutting down photosynthesis). Smaller objects cause regional damage but would be unlikely to precipitate a major extinction like the Cretaceous–Tertiary event.

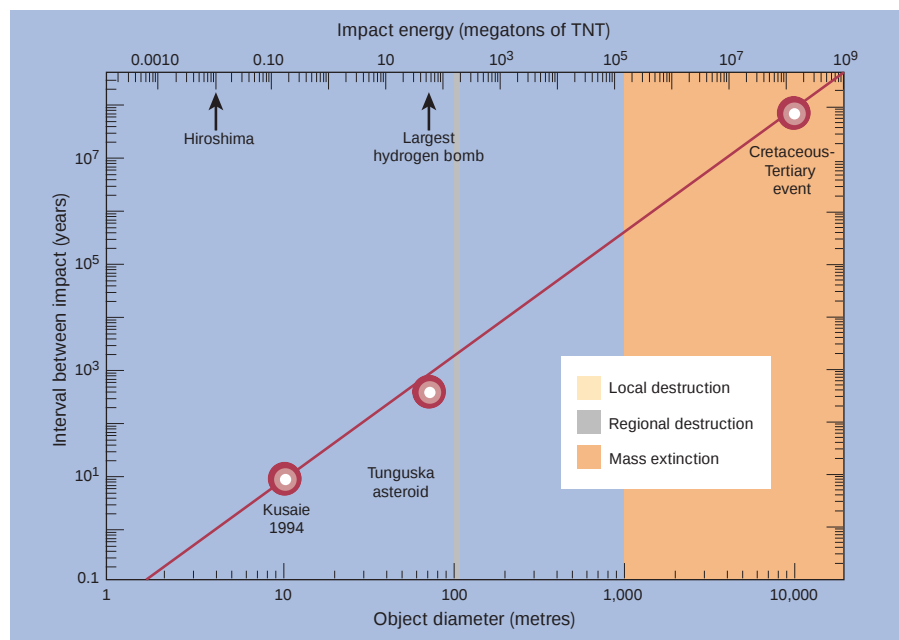


Figure 1 Previous impacts by near-Earth objects (NEOs). Three well-documented impacts are marked. The dinosaur-killing Cretaceous–Tertiary event was caused by an asteroid 10 km in diameter. Such objects hit the Earth about once every 100 million years. A 50-metre object similar to that which destroyed the Tunguska pine forest strikes the Earth once every few centuries. Atmospheric explosions observed by satellites above the South Pacific island of Kusaie in 1994 might occur every decade. NASA is committed to discovering 90% of all NEOs larger than 1 km in diameter (energy, 100,000 megatons). These objects are capable of wreaking global havoc but are very rare. The vast majority of the dangerous objects (those less than 1 km but more than ~100 metres in diameter) have yet to be discovered.

Last summer, astronomers devised a new risk-assessment scale, similar to the Richter scale used for earthquakes, to help the public understand the hazard posed by a given NEO. The so-called Torino scale ranges from zero (no chance of a collision) to 10 (certain collision causing global devastation). No known NEO has yet had a Torino number greater than one. This is just as well because we presently have no coherent plan of action should a real threat arise. The simplest option — massive evacuation of the impact site — would be impractical because of the positional uncertainties and large numbers of people involved, and would be ineffective because the damage from large NEOs will be global. One option that has been discussed is the thermonuclear destruction of the incoming NEO (a bad idea because the shower of debris produced by the exploding NEO might be as damaging as the initial object, and would be radioactive). Given enough time, the NEO might be deflected from an Earth-intersecting path by a series of smaller explosions, or by attaching rockets or solar sails that use radiation pressure from the Sun.

The focus on NEOs larger than 1 km ignores the threat from smaller but much more numerous objects. The Earth's atmosphere offers little protection against objects larger than 100 metres in diameter⁴. These smaller objects outnumber NEOs larger than 1 km by a factor of 100, so they are much

more likely to strike in our lifetimes. There is a 1% chance that the Earth will be struck by a 300-metre NEO in the next century⁴. Such an impact would deliver a withering 1,000-megaton explosion and cause perhaps 100,000 deaths. If the impact occurred in or near a densely populated region — the eastern seaboard of the United States, for instance, or Western Europe or coastal Asia — the fatalities could easily rise into the tens of millions.

Neither can we take refuge in the fact that 70% of the Earth is covered by oceans. Impact-induced tsunamis could wipe out coastal cities over a wide area. So, to have practical value, surveys should not be limited to the (observationally easy but numerically rare) 1-km NEOs, but should instead catalogue objects at least down to the few-hundred-metre size range⁵. What is needed is a more ambitious survey to completely identify the population of small, potentially threatening NEOs.

The strategy for such a survey has been explored by Alan Harris of the Jet Propulsion Laboratory⁶. He argues that the whole sky must be surveyed on a monthly basis with a sensitivity about 100 times greater than current NASA-sponsored surveys. How can this be done? A large (6–8-metre) telescope is required, with a wide field of view tiled with CCD (charge-coupled device) optical detectors and connected to a massive computer array capable of meeting the huge

data-processing demands. The technology exists and tentative designs are beginning to appear⁷⁻⁹. Such a telescope, which would have many applications in other branches of astronomy, is projected to cost about \$100 million (about half the price of a Jumbo jet). What is missing is any sign that such a facility will be funded by governments and their agencies. Perhaps astronomers can attract the interest of private donors in the search for threatening NEOs. If not, it seems we will have to face the asteroidal impact hazard with our eyes wide shut.

David Jewitt is at the Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive,

Honolulu, Hawaii 96822, USA.

e-mail: jewitt@ifa.hawaii.edu

1. Rabinowitz, D., Helin, E., Lawrence, K. & Prado, S. *Nature* **403**, 165–166 (2000).
2. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. *Science* **208**, 1095–1108 (1980).
3. Morrison, D. *The Spaceguard Survey: Report of the NASA International Near-Earth Object Detection Workshop* (Jet Propulsion Laboratory, Pasadena, 1992).
4. Chapman, C. R. & Morrison, D. *Nature* **367**, 33–40 (1994).
5. Binzel, R. P. *et al.* *From the Pragmatic to the Fundamental: The Scientific Case for Near-Earth Object Surveys* (1999).
6. Harris, A. *Planet. Space Sci.* **46**, 283–290 (1998).
7. <http://www.dntlescope.org>
8. <http://irtf.ifa.hawaii.edu/NPT/npt.html>
9. <http://www.rc-obs-azur.fr/schmidt/general/NEOSurvey.html>

Plant development

Gateway to the chloroplast

Kenneth Cline

Chloroplasts are essential organelles in plants and algae. In plants, they are formed by differentiation of a progenitor plastid, the proplastid. This process involves the import of roughly 2,000 different proteins from the cytosol, across the outer and inner membranes of the chloroplast envelope¹. Proteins destined for the interior of the chloroplast are made as precursors, and the amino-terminal 'transit peptides' that direct import are removed once the protein is inside. *In vitro* studies have helped us to understand how chloroplasts import proteins¹. However, models based on these studies have not been examined *in vivo*.

Bauer *et al.*² have now addressed the *in vivo* function of the putative import receptor Toc159. (The name Toc refers to the translocation apparatus of the outer-envelope membrane.) On page 203 of this issue they report the identification of a mutant *Arabidopsis thaliana*, known as *ppi2*, that lacks the Toc159 protein. The proplastids of *ppi2* fail to develop into chloroplasts and the plants die at the seedling stage, suggesting that Toc159 is essential for chloroplast development. The lethality of *ppi2* was unexpected because, in other protein-translocation systems, deletion of single receptors is generally not lethal³. Perhaps even more exciting is the fact that this study has identified two additional import receptor proteins (Toc120 and Toc132), and this may be the key to unravelling how plastids recognize a strikingly wide range of transit peptides.

Chloroplasts evolved from a photosynthetic bacterium that was taken up by a host cell. During evolution, most of the bacterial genes were transferred to the host's nucleus, so a mechanism was needed to return the corresponding proteins to the chloroplast. Successfully transferred genes acquired

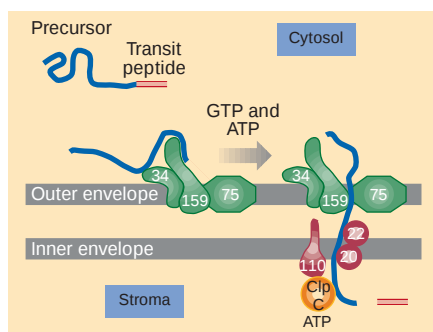


Figure 1 Model for protein import across the outer and inner chloroplast envelope membranes. A Toc complex of Toc159, Toc75 and Toc34 has been identified by *in vitro* studies, and Bauer *et al.*² have now found two additional complexes that contain Toc132 and Toc120, respectively, in place of Toc159. Component proteins in the inner membrane are referred to as the Tic apparatus; for example, Tic110, Tic22 and Tic20. ClpC is a stromal chaperone that may provide the driving force for translocation.

coding sequences for peptides that target proteins to the chloroplast, and the evolving chloroplast developed machinery to import these precursor proteins¹. Because we can reconstitute import into chloroplasts *in vitro* (Fig. 1), various steps of the import process have been dissected and components of the machinery responsible have been isolated.

The identified components, although mostly hitherto-unknown proteins, seem to perform functions common among protein-translocation systems. For example, Toc159, which is found in the outer-envelope membrane, is considered to be a receptor because it interacts directly with the transit peptides of at least two precursors⁴, and because antibodies to Toc159 inhibit precursor binding and import⁵. Another protein called Toc75 seems to form at least part of the channel

through which proteins cross the outer-envelope membrane⁶. The chaperone ClpC, found in the interior (or stroma) of the chloroplast, may be the ATP-powered motor that 'pulls' precursor proteins through the channels¹.

The diversity of transit peptides promises to make characterization of receptor function in the chloroplast-import system particularly challenging. Transit peptides vary in size from 30 to more than 100 residues, they lack consensus sequences that could serve as address tags, and they have no apparent 'signature' secondary structures. Rather, transit peptides share only compositional features, such as a deficiency of acidic amino acids and an abundance of serine and threonine. The apparent degeneracy of transit peptides is probably due to their origins, which must have been a random process in which the organellar genes inserted themselves into nuclear DNA in-frame with bits of other coding regions.

In this context, the finding by Bauer *et al.*² that there are at least three import receptors — Toc159, Toc132 and Toc120 — is highly significant. These proteins are all expressed in chloroplast-containing tissue, they have similar structures, are orientated similarly in the membrane, and are present in outer-envelope translocation complexes. They are unlikely to be functionally redundant for at least two reasons. The first is the severity of the *ppi2* mutation compared with, for example, *ppi1* (another chloroplast-import mutant in which a member of the Toc34 family is disrupted⁷). Whereas the *ppi2* mutation is lethal, *ppi1* plants show delayed chloroplast development but they do survive and eventually appear similar to wild-type plants. Second, although Toc159, Toc132 and Toc120 have highly conserved carboxy-terminal GTP-binding and membrane-anchoring domains, their large cytosol-exposed domains — that is, those domains likely to make initial contact with transit peptides — show considerable sequence divergence.

A more attractive model is that different receptors, with different (possibly overlapping) specificities, accommodate different classes of precursor proteins. Bauer *et al.* suggest that Toc159 may be dedicated to photosynthetic proteins, whereas Toc132 and Toc120 could accommodate non-photosynthetic proteins. The chloroplast belongs to a family of developmentally inter-related organelles collectively called plastids. Plants tailor the type of plastid in each organ by expressing different sets of genes, depending on their metabolic needs. Many proteins are common among plastids, whereas others are specific to a particular plastid. So, a plausible idea is that different types of plastids, such as amyloplasts or chromoplasts, possess different receptor combinations. In support of this idea, transit

peptides of certain proteins determine the efficiency of precursor import into various types of plastid^{8,9}.

How can we improve our understanding of the way chloroplasts recognize precursor proteins? The possibilities described above make specific predictions that could be tested *in vivo* with the methodologies developed by Bauer *et al.*² and Jarvis *et al.*⁷. The genome sequence of *Arabidopsis* will be completed later this year, allowing a definitive assessment of the number of *Toc* and *Tic* (translocation apparatus of the inner-envelope membrane) genes. There are several large-scale projects for identifying *Arabidopsis* plants with disruptions of specific genes, so we should soon have disruptions in each of the receptor genes. *Arabidopsis* can be transformed with both wild-type and modified genes under various expression regimes, allowing one receptor to be substituted for another. By examining different

organs of the resulting plants, it should be possible to detect defects in development of the various plastid types. Finally, combined with *in vitro* assays using the isolated plastids⁷, this approach could be used to match each receptor with its set of substrate transit peptides. ■

Kenneth Cline is in the Department of Horticultural Sciences, University of Florida, 1137 Fifield, Gainesville, Florida 32611, USA.

e-mail: kcline@ufl.edu

1. Keestra, K. & Cline, K. *Plant Cell* **11**, 557–570 (1999).
2. Bauer, J. *et al.* *Nature* **403**, 203–207 (2000).
3. Schatz, G. & Dobberstein, B. *Science* **271**, 1519–1526 (1996).
4. Kouranov, A. & Schnell, D. J. *J. Cell Biol.* **139**, 1677–1685 (1997).
5. Hirsch, S., Muckel, E., Heemeyer, F., von Heijne, G. & Soll, J. *Science* **266**, 1989–1992 (1994).
6. Hinnah, S. C., Hill, K., Wagner, R., Schlicher, T. & Soll, J. *EMBO J.* **16**, 7351–7360 (1997).
7. Jarvis, P. *et al.* *Science* **282**, 100–103 (1998).
8. Wan, J., Blakely, S. D., Dennis, D. T. & Ko, K. J. *Biol. Chem.* **271**, 31227–31233 (1996).
9. Reinbothe, C., Lebedev, N., Apel, K. & Reinbothe, S. *Proc. Natl Acad. Sci. USA* **94**, 8890–8894 (1997).

Telomeres

A tale of ends

Victoria Lundblad

During the past few years in the field of chromosome biology, much attention has focused on the very tip of the chromosome — the telomere. This is because functional telomeres are essential for continuous cellular proliferation, an observation that has profound implications for our understanding of ageing and cancer. Studies of telomere biology have been aided by the discovery of numerous genes required for telomere replication in ciliates, yeast, mice and humans. But missing from this genetic pantheon has been the nematode worm *Caenorhabditis elegans*, one of the premier organisms for gene identification and analysis. Now, however, on page 159 of this issue, Ahmed and Hodgkin¹ describe an elegant screen for *C. elegans* mutants with mortal germ lines. The first gene to be characterized from this work, *mortal germline-2* (*mrt-2*), plays double duty in the nucleus — it is required not only for telomere replication, but also for monitoring damaged DNA.

The strategy used to identify the *mrt-2* gene was based, in part, on what we know about telomere dysfunction in other systems. For growth of single-celled eukaryotes such as ciliates and yeast, continuous telomere replication is needed. Loss of telomerase — the enzyme responsible for replicating telomeres — results in a cell line that can be maintained for only a limited number of cell divisions, whereas restoration of telomerase function confers unlimited cellular propagation^{2,3}. In mammals, by contrast, telomere replication is highly

regulated, and many tissues do not sustain high levels of telomerase. However, the germ line, which is passed from one next generation to the next, is an immortal cell lineage that must be able to proliferate indefinitely. Here, as expected, telomeres are maintained at their full length⁴, and, if telomerase is

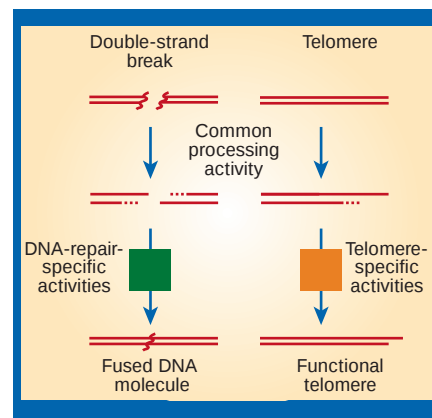


Figure 1 Proposed set of common events at DNA double-strand breaks and telomeres. Evidence is growing for the idea that these two pathways are partially linked. Ahmed and Hodgkin¹ show that the *mortal germline-2* (*mrt-2*) gene, which is required for telomere replication, is also involved in monitoring DNA for double-strand breaks. The idea is that processing to resect one strand of a telomere or double-strand break could be resolved by activities that are specific for telomere maintenance (such as telomerase) or DNA repair (such as a DNA ligase).

eliminated by a directed knockout in mice, the result is eventual sterility⁵.

Ahmed and Hodgkin¹ screened for mutants of *C. elegans* with mortal germ lines — and, hence, potential deficiencies in telomere replication — by identifying mutant lines of worms that could produce offspring initially but eventually became sterile. To do so they exploited a unique feature of *C. elegans* biology: as a self-fertilizing hermaphrodite, a single worm can give rise to a pedigree of genetically identical descendants. The authors examined successive brood sizes of 400 mutagenized lines of worms and identified 16 *mrt* mutants with ‘delayed’ sterility (that is, the number of offspring declined with each successive generation). This progressive defect is reminiscent of the steady decline in cell proliferation shown by yeast strains deficient for telomere replication^{2,3}. Indeed, genomes of *mrt-2* worms show a steady erosion of telomeric DNA as well as a high frequency of late-onset end-to-end chromosome fusions, both of which characterize telomerase-deficient fission yeast and mouse mutants^{3,6}.

Because the *mrt-2* mutant has such a profound telomere-replication defect, could it be a component of telomerase? Surprisingly it is not. Instead, the *mrt-2* gene is the *C. elegans* homologue of a checkpoint gene, conserved from yeast to humans, that is involved in the response to DNA damage. The sequence of *mrt-2* also explains another property of the mutant worms: they are more sensitive than normal worms to agents that introduce double-strand breaks in DNA, such as X-rays. This would be expected if the ability to sense — and subsequently repair — DNA damage is altered. Mutation of several checkpoint genes in other organisms, including the fission yeast homologue of the *mrt-2* gene⁷, has relatively modest consequences for telomere maintenance. So the *mrt-2* mutant stands out for the severity of its defects in both telomere replication and the DNA-damage response.

This is not the first time that telomeres and double-strand breaks have shared company, because several proteins are known to be required for both telomere function and DNA repair⁸. These results challenge the simple, yet conceptually appealing, idea that telomeres and double-strand breaks are functionally distinct termini⁹. This concept has driven much of the thinking about chromosome end protection over the past several decades. Telomeres must avoid fusion with other termini, whereas successful repair of DNA breaks demands re-joining of the broken ends. The realization that certain DNA-repair proteins are also required for normal telomere function raises the possibility that chromosome ends may transiently resemble a double-strand break at some point during the cell cycle. However, completion of the full sequence of events that result in end-

joining clearly does not occur at telomeres.

If this speculative model for the connection between telomeres and damaged DNA is correct, might the MRT-2 protein facilitate telomere function as a direct consequence of its checkpoint activity? Ahmed and Hodgkin¹ suggest that the telomere must be recognized as a double-strand break by checkpoint proteins before telomerase can act on it. For instance, MRT-2 (and its counterparts in other organisms) could contribute to strand-specific processing of the telomere, which is thought to be essential in maintaining chromosome ends¹⁰. Similar processing at double-strand breaks, also activated by checkpoint proteins¹¹, would be followed by a different set of events, leading instead to fusion of broken ends (Fig. 1).

Such a mechanism suggests that the cell uses an elegant economy when it handles DNA termini. However, we still do not know

what activity dictates whether a processed end becomes a telomere or a fused DNA molecule. So the telomere puzzle — which was presented to us by Barbara McClintock almost 60 years ago⁹ — has yet to be completely solved. ■

Victoria Lundblad is in the Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

e-mail: lundblad@bcm.tcm.edu

1. Ahmed, S. & Hodgkin, J. *Nature* **403**, 159–164 (2000).
2. Lundblad, V. & Szostak, J. W. *Cell* **57**, 633–643 (1989).
3. Nakamura, T. M., Cooper, J. P. & Cech, T. R. *Science* **282**, 493–496 (1998).
4. Cooke, H. J. & Smith, B. A. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 213–219 (1986).
5. Lee, H.-W. *et al.* *Nature* **392**, 569–574 (1998).
6. Blasco, M. A. *et al.* *Cell* **91**, 25–34 (1997).
7. Dahlen, M. *et al.* *Mol. Biol. Cell* **9**, 611–621 (1998).
8. Shore, D. *Science* **281**, 1818–1819 (1998).
9. McClintock, B. *Genetics* **26**, 234–282 (1941).
10. Wellinger, R. J., Ethier, K., Labrecque, P. & Zakari, V. A. *Cell* **85**, 423–433 (1996).
11. Lydall, D. & Weinert, T. *Science* **270**, 1488–1491 (1995).

Phase transitions

Jumping between liquid states

Paul McMillan

Crystalline solids can exist in different forms — polymorphs — with different structures and bonding patterns. Such polymorphs are usually stable under different conditions. For example, carbon occurs naturally as two different polymorphs, graphite and diamond, which are stable at different pressures and temperatures. The transformation between graphite and diamond is classified thermodynamically as a first-order phase transition, in that the structure and energy change abruptly at the transition. First-order phase transitions in non-crystalline materials — for example, the abrupt condensation of a gas to a liquid — are also well known.

The existence of liquid (rather than solid) polymorphs is only just beginning to be recognized. For a long time, the idea of a first-order transition between two liquids was not considered seriously, because of the view that liquids have rapidly changing structures that vary smoothly with temperature and pressure. Growing evidence for first-order transitions in liquids has come from analysis of their physical properties and from computer simulations^{1,2}. On page 170 of this issue, Katayama *et al.*³ describe a direct structural study of liquid phosphorus under high pressure, using synchrotron X-ray diffraction. Simply by increasing the pressure, the structure of liquid phosphorus jumps suddenly from a relatively open molecular structure to a new polymeric form with higher density.

To place the results of Katayama *et al.* in context we need to understand the melting

behaviour of certain liquids. Crystals are normally denser than their corresponding melts, so the melting temperature increases with increasing pressure (that is, the melting slope is positive). Important exceptions to this rule include silicon and water ice, which have a negative melting slope. Several other crystalline elements and compounds show maxima in their melting curves as the pressure increases⁴.

Such unusual behaviour is understood in terms of a 'two-state model', in which low- and high-density states (associated with different atomic arrangements) coexist within the liquid. As pressure is applied, the relative proportion of high- to low-density states increases. The result is that the liquid density at high pressure can exceed that of the crystalline state, and the slope of the melting curve becomes negative. Thermodynamic analysis shows that the two-state behaviour passes into two-phase behaviour below a certain temperature^{1,4}, in so far as the two phases have different thermodynamic properties, such as energy or entropy (the degree of disorder). So by applying pressure to the liquid below this critical temperature, a first-order transition between two different liquid phases is expected to occur.

The case of water is interesting. Experiments reveal a transition between low- and high-density forms of amorphous ice, leading to the suggestion that a pressure-driven phase change might occur in the liquid^{2,4,5}. 'Normal' water corresponds to the high-density liquid, which explains why ice floats on it. The low-density, highly viscous liquid has yet to be observed experimentally, but it exists in computer simulations. Other pressure-induced transformations between low- and high-density liquids have been seen in glassy systems that have an open structure (for example, SiO_2)^{1–4}. But glasses are not in thermodynamic equilibrium, so such transformations do not correspond to true phase transitions from one stable liquid to another. A second liquid phase has been seen in supercooled molten $\text{Y}_2\text{O}_3\text{-Al}_2\text{O}_3$ (ref. 6). But two phases had not been directly observed in a stable

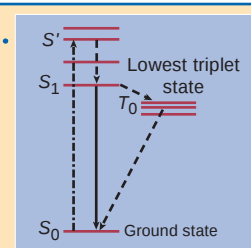
Nuclear polarization

More spins for protons

Protons have an intrinsic spin or angular momentum that becomes important when the proton interacts with other polarized particles or with an electromagnetic field. Particle physics experiments with polarized protons require high magnetic fields (2.5 to 5 tesla) and extremely low temperatures (0.3 to 1 K) to reach 70% polarization. In contrast, magnetic resonance imaging techniques that operate at room temperature and magnetic fields of 1 tesla only achieve 0.0003% polarization. The production of polarized protons in a modest magnetic field (0.3 tesla) at liquid-nitrogen temperatures

(77 K) has now been reported by Iinuma *et al.* (*Phys. Rev. Lett.* **84**, 171–174; 1999).

Iinuma *et al.* achieve high levels of proton polarization (up to 32%) by doping their materials with a small amount of pentacene. Electrons in pentacene can be excited to higher states (S' and S_1 in the figure) with a laser beam. Around 2% of the electrons transfer from S_1 to the lowest triplet state T_0 . Using microwaves tuned to the frequencies of the electron and proton spins, the electron polarization is transferred to nearby protons. Electrons in the triplet state decay into the ground state, where the proton



spin remains polarized.

Early experiments at liquid-nitrogen temperatures polarized just 13% of the protons (the theoretical limit is 73%). Iinuma *et al.* trebled this using a powerful pulsed dye laser (rather than a nitrogen laser) to optically excite the electrons to the first excited state of pentacene. The authors suggest that their technique might enhance the sensitivity of NMR at higher temperatures as well as having an impact on particle physics. Sarah Tomlin

simple liquid until the work of Katayama *et al.*³.

Previously, Brazhkin and colleagues observed large changes in the electrical conductivity of simple liquids (iodine, sulphur and selenium) at high pressure, which were attributed to liquid–liquid phase transitions^{4,7}. But these changes could equally well occur in a two-state regime within a single liquid phase⁴. Katayama *et al.*³ provide a method to determine which type of thermodynamic behaviour is associated with changes in liquid properties as the pressure varies. They use synchrotron X-ray diffraction methods to study directly the structure of liquid phosphorus at high pressures and temperatures. They observe at a pressure near one gigapascal, a sharp transition between the molecular liquid present at atmospheric pressure and a high-density polymeric form. The two forms coexist within a narrow pressure range, as determined by the composite X-ray spectrum, and the transition could be reversed by lowering the pressure — providing direct evidence for a pressure-driven structural transition in the liquid above the melting point.

The evidence now available indicates that first-order phase transitions do occur in the liquid state, and that the phenomenon is probably widespread^{1–4}. We must revise our conventional picture of liquids as entities with continuously varying averaged structure. Instead, each liquid is constrained by distinct ‘configurational landscapes’ that are explored within a given range of density (or pressure)⁸. Liquid polymorphs with different structures and physical properties have landscapes that are separated by high-energy

barriers that determine the first-order nature of liquid–liquid transitions². This ‘configurational landscape’ model may even be applied to biomolecules, in which transformations between unfolded and folded structures of proteins may be modelled by analogous phase transitions⁸.

Multiple liquid–liquid transitions may occur for a given substance as the pressure is varied^{1,4,7}. At the lowest density, the liquid–gas transition is then the ‘final’ member of a suite of such transitions. The experiments of Katayama *et al.*³ point the way not only to structural investigations of liquid polymorphs, but also to gaining a broader understanding of relationships between liquids and gases, by providing direct confirmation that pressure-driven transitions between different liquid phases do occur. An even more intriguing possibility is that first-order transitions may occur between gases. A ‘normal’ gas is already recognized as distinct from an ionized plasma. Perhaps transitions between non-ionized gaseous phases might occur at extremely low temperatures and at low densities.

Paul McMillan is at the Center for Solid State Science, Arizona State University, Tempe, Arizona 85287-1604, USA.

e-mail: pmcmillan@asu.edu

1. Poole, P. H., Grande, T., Angell, C. A. & McMillan, P. F. *Science* **275**, 322–323 (1997).
2. Angell, C. A. *Science* **267**, 1924–1935 (1995).
3. Katayama, Y. *et al. Nature* **403**, 170–173 (2000).
4. Ponyatovsky, E. G. & Barkalov, O. I. *Mater. Sci. Rep.* **8**, 147–191 (1992).
5. Aasland, S. & McMillan, P. F. *Nature* **369**, 633–636 (1994).
6. Mishima, O. & Stanley, H. E. *Nature* **396**, 329–335 (1998).
7. Brazhkin, V. V., Popova, S. V. & Voloshin, R. N. *High Pressure Res.* **15**, 267–305 (1997).
8. Angell, C. A. *Physica D* **107**, 122–142 (1997).

Palaeontology

Fossil fish up for election

Meemann Chang

The bony fishes (osteichthyans) were the progenitors of tetrapods, without which none of us would be here. So we have good reason to be interested in bony-fish evolutionary history. The group is divided into the ray-finned and lobe-finned fishes, and recently discovered early bony-fish specimens have a mixture of characters of each type. Moreover, these specimens show some features formerly thought to be present only in the more primitive cartilaginous fishes (chondrichthyans — sharks and rays) and an extinct group called placoderms. So palaeontologists have been puzzling over which group the new forms should be assigned to, and how all of the various groups are related to each other.

On page 185 of this issue¹, Basden *et al.* report on another bony-fish fossil that

comes from southeastern Australia and is designated AM101607. It looks ray-fine-like, but it also exhibits some characters previously found only in lobe-finned fishes and others found only in chondrichthyans and placoderms. Last year, Zhu *et al.*² discussed another fish of this kind, *Psarolepis*, from China and Vietnam, which had been earlier described by Yu³ and is more lobe-fine-like in appearance. Both *Psarolepis* and AM101607 are from the Early Devonian period, about 400 million years ago, and so are among the oldest bony fishes known. The respective authors^{1,2} each propose that the specimen they describe is a candidate to be a member of the basal group (the ‘ancestor’) of bony fishes, but Basden *et al.*¹ consider that their fish is the more promising contender.

Specimen AMF101607 consists only of a

large portion of a braincase, but its three-dimensional preservation reveals many anatomical details. In several respects it resembles two well-preserved ray-finned fishes, *Mimia* and *Moythomasia*, from the Late Devonian (about 350 million years ago) of western Australia⁴; similarities lie in the ornamentation of the dermal bones; in the pattern of the skull roof (Fig. 1a, b), and of the sensory canal and pit line; and in the structure of the lateral wall of the orbito-temporal region of the braincase. It also has some characters in common with *Psarolepis* and other lobe-finned fishes. Other characters, however, are shared with ray-finned fishes and chondrichthyans, rather than lobe-finned fishes (the postorbital process; the broad spiracular groove; and the lateral commissure joining the base of the orbito-temporal region of the braincase). But, most strikingly, AMF101607 has some features never found in either lobe-finned or ray-finned fishes, such as an eyestalk.

Eyestalks are rod-like structures that grow out of the inner wall of the orbit and have a flat, rounded knob at the end which abuts against the eyeball⁵. They were thought to be unique to chondrichthyans and placoderms, where they (and the endocranium as a whole) are cartilaginous and can be seen in full only in recent sharks and rays (Fig. 2a). In placoderms, now represented only by fossils, the cartilaginous endocranium is frequently lined with a thin layer of calcified cartilage. In these fishes, the eyestalk itself is not preserved. Even its base, which is part of the endocranium, has been found in only a few taxa, such as *Jagorina* and *Macropetalichthys*⁶, where it is evident as a stub (Fig. 2b). In most cases there is only a hole in the inner wall of the orbit at the site for eyestalk attachment⁷ (Fig. 2c).

It seems that the minerals filling the base of the eyestalk in Basden and colleagues’ specimen were etched out during acid preparation of the fossil, and there is only an opening left in the inner wall of the orbit. Yet the rim of the opening is clearly everted,

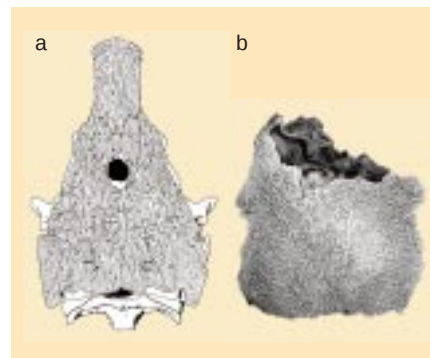


Figure 1 Similar patterning of the skull roof of Devonian fossil ray-finned fishes from Australia. a, *Mimia*, from the Late Devonian⁴. b, The new specimen AMF101607 from the Early Devonian, as described by Basden *et al.*¹.

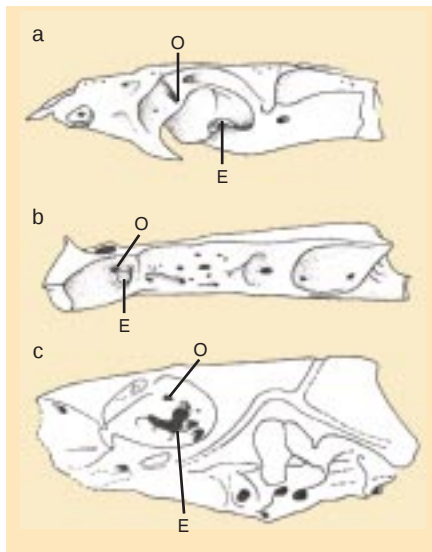


Figure 2 Fish braincases in lateral view, showing the eyestalk, or its base or site of attachment (all marked as E; see also Fig. 2 on page 186). The eyestalk is a diagnostic feature that was thought to be unique to chondrichthyans and placoderms, but is evident in Basden and colleagues' fish¹. a, *Chlamydoselachus* (chondrichthyan), a recent shark⁵. b, *Jagorina*, a placoderm⁶. c, *Brindabellaspis*, a placoderm¹. O, the opening for the optic nerve.

matching the condition in *Jagorina*. So there is no doubt that AMF101607 had an eyestalk. In consequence, eyestalks can no longer be considered to be unique to chondrichthyans and placoderms.

This is not a trivial point. Changing knowledge about the distribution of key characteristics often radically alters our understanding of the evolutionary relationships of their bearers. Take feathers, for example — as soon as it was found that theropod dinosaurs possessed them, feathers could no longer be regarded as unique to birds. The revelation that the new fish from Australia had an eyestalk has similar significance (if less public appeal).

Based on the cranial features of AMF-101607 and of examples of other groups in the evolutionary tree constructed by Zhu *et al.*², and using a computer program called PAUP, Basden *et al.* produced several alternative positions of their fish and *Psarolepis* in the tree (see Fig. 3 on page 187). Their preference is for one with the new form at the base of the bony fishes and *Psarolepis* at the base of the lobe-finned fishes.

Uncertainty about the positions of the two forms in Basden and colleagues' tree arises mainly from insufficient information. Specimen AMF101607 is very similar to the two Late Devonian ray-finned fishes, so we ought to compare them more thoroughly. The eyestalk-attachment area in AMF-101607 seems to occupy exactly the same position as the opening for the optic nerve in *Mimia* and *Moythomasia* (and is almost

the same shape and size), so it could be that these fishes also had an eyestalk. Comparison of the new form with *Mimia* and *Moythomasia* will help to identify the closest relative of bony fishes, and may even tell us something about the evolution of jawed vertebrates as a whole.

Other new knowledge also has to be taken into account. The characters previously held to be unique to osteichthyans have recently been identified in a chondrichthyan from Bolivia⁸, and a molecular study places the cartilaginous fishes within the bony-fish evolutionary tree⁹. All of these developments challenge our understanding of the relationships among the major groups of primitive back-boned animals and demand a profound change of conventional wisdom. The new form reported by Basden *et al.* is a

welcome addition to the list of candidates for the position of 'ancestor' of the bony fishes. But whether it will be elected remains to be seen.

Meemann Chang is at the Institute of Vertebrate Paleontology and Paleoanthropology, PO Box 643, 100044 Beijing, China.

e-mail zhangmm@public.bta.net.cn

1. Basden, A. M., Young, G. C., Coates, M. I. & Ritchie, A. *Nature* **403**, 185–188 (2000).
2. Zhu, M., Yu, X. & Janvier, P. *Nature* **397**, 607–610 (1999).
3. Yu, X. *J. Vert. Paleontol.* **18**, 261–274 (1998).
4. Gardiner, B. G. *Bull. Brit. Mus. Nat. Hist. (Geol.)* **37**, 173–428 (1984).
5. Allis, E. P. *Acta Zool.* **4**, 123–221 (1923).
6. Stensiö, E. in *Traité de Paléontologie IV*, Vol. 2 (ed. Piveteau, J.) 71–692 (Masson, Paris, 1969).
7. Young, G. *Palaeontographica* **167A**, 10–76 (1980).
8. Janvier, P. *Trends Ecol. Evol.* **14**, 298–299 (1999).
9. Rasmussen, A.-S. & Arnason, U. *J. Mol. Evol.* **48**, 118–123 (1999).

Neurobiology

Internal model visualized

Masao Ito

How can we instantly recognize a familiar object? Probably because, in the brain, we already have a model of that object which is activated through vision. Similarly, we can quickly comprehend what we hear because the brain is likely to contain a model representing the meanings of the sounds we encounter. And we can probably carry out complex movements so easily and accurately because the cerebellum provides a model of what is to be moved. Such internal models provide an attractive explanation for the brain's subtle cognitive and control mechanisms, but there has been no way to investigate them experimentally. Now, however, on page 192 of this issue, Imamizu *et al.*¹ report a set of brain-imaging data that

provide the first evidence for such a model being formed in the cerebellum.

The basis of the work done by Imamizu *et al.* is the unique theory² of the adaptive-control system with two degrees of freedom (Fig. 1). It works like this. When we move a hand, a desirable movement worked out somewhere in the brain is conveyed as an instruction to the motor area and its related regions in the cerebral cortex. These areas in turn generate command signals, which act on the hand/arm system to carry out the movement. Information about the realized movement is conveyed through the visual system back to the motor and related areas, and compared with the movement received as an instruction. This feedback ensures that

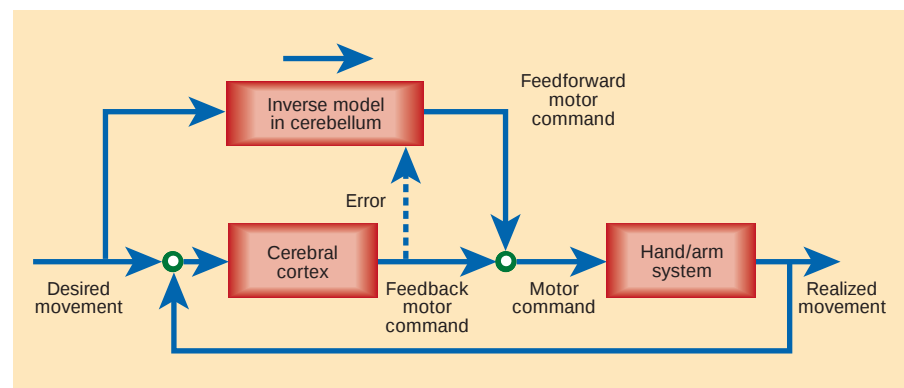


Figure 1 Block diagram of the two degrees of freedom adaptive-control system with feedback-error learning mechanism. This system combines a feedback control by the cerebral cortex and a feedforward control by the cerebellum. The cerebral cortex compares the instructed, desired movement with a realized movement by sensory feedback, whereas the cerebellum receives only the instruction. To realize a desired movement without feedback of the realized movement, the cerebellum needs to form an inverse model of the hand/arm system, as visualized by Imamizu *et al.*¹ using brain-imaging data. (Modified from ref. 2.)

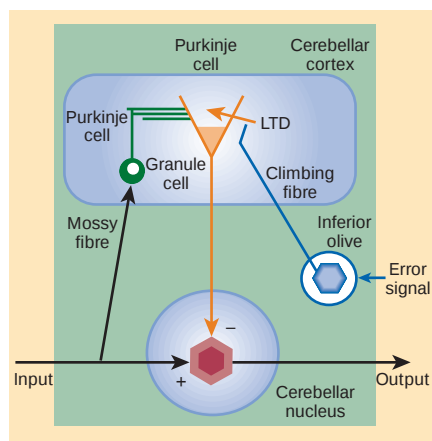


Figure 2 Basic structure of the neuronal circuitry in the cerebellum. A single functional 'unit' is shown. Mossy-fibre signals provide inputs to the unit and neurons in the cerebellar nucleus generate outputs. Climbing-fibre signals represent errors, reorganizing internal connections within the unit to modify its input–output relationship. LTD, long-term depression. (Adapted from ref. 3.)

the realized movement is similar to the 'instruction' movement. The instruction may also be converted to command signals by the cerebellum, instead of by the cerebral cortex.

Even without such feedback, this pathway could produce a movement that closely matches the movement received as an instruction. For this to happen, however, the cerebellum would need to contain the inverse of the dynamics or kinematics of the hand/arm system — in other words, the cerebellum would need to contain an inverse model of the hand/arm system. The model is formed and updated through learning, by referring to the errors between the intended and performed movements, as detected by the motor and related areas of the cerebral cortex (feedback-error learning)².

To test this hypothetical prediction, Imamizu *et al.*¹ asked a human subject lying in a functional magnetic resonance imaging (fMRI) scanner to manipulate a computer mouse. During the so-called 'baseline' period, the subject followed a moving square target with a cross-hair cursor on a screen. Then, during the test, the position of the cursor was rotated 120° around the centre of the screen to provide a new mouse 'condition'. At first, during the test period, large regions of the cerebellum were significantly activated compared with their activity during the baseline period. But this activation decreased after repeated test trials, in parallel with a reduction of the tracking errors. However, certain restricted sub-regions of the brain (near the posterior superior fissure) continued to be activated. The authors propose that this remaining activity represents an internal model that is formed during the repeated test trials. This model defines the

new relationship between movement of the cursor and of the mouse.

The internal-model concept for motor learning is reasonable if we consider the known cellular mechanisms of the cerebellum. The cerebellum has a compartmentalized structure that consists of numerous small, functional units, each of which contains a neuronal network, composed of Purkinje and other neurons, organized in a geometrically beautiful way (Fig. 2). Each unit can change its input–output relationship through learning, which is driven by error signals conveyed to the Purkinje cells by a unique input structure, the climbing fibres. Climbing-fibre signals induce long-term depression (LTD) of transmission from a major input to the Purkinje cells, via the mossy fibre–parallel fibre pathway³. LTD depresses those synapses between the parallel fibres and Purkinje cells that are involved in making errors, thereby reorganizing internal connections towards a reduction of the errors. Through this learning mechanism a cerebellar unit could modify its input–output relationship until it represents the model required for precise control.

A crucial question is whether an increase in local blood flow in the cerebellum, as detected by fMRI, really represents the formation and maintenance of an internal model. If LTD is the only synaptic plasticity that underlies the learning mechanism of the cerebellar circuitry, the learning is unlikely to be accompanied by an increase in electric impulse discharges (which may be reflected as increased local blood flow). So the authors assume that excitatory and inhibitory synaptic transmissions to the Purkinje cells are facilitated, and that electric-impulse discharges of the Purkinje cells increase.

But in the complex neuronal circuitry of the brain there may be other possibilities. For example, many chemical reactions underlie the induction of LTD, including the release of nitric oxide, which has a well-known action of relaxing blood capillaries. The next question to address experimentally is how not only the electric-impulse discharges, but also the complex chemical processes, contribute to increases in local blood flow in the cerebellum.

As neuroscientists strive to understand the molecular and cellular events that occur in neurons, effective technologies and methodologies for experimentally investigating computational and cognitive principles of the brain are still in short supply. The success of Imamizu *et al.*¹ in visualizing an internal model in the cerebellum should therefore provide tremendous encouragement for other researchers in the field. ■

Masao Ito is at the Brain Science Institute, RIKEN, Wako, Saitama 351-0198, Japan.

e-mail: masao@mailsv.brain.riken.go.jp

- Imamizu, H. *et al.* *Nature* **403**, 192–195 (2000).
- Kawato, M. & Gomi, H. *Biol. Cybern.* **68**, 95–103 (1992).
- Ito, M. *Trends Cog. Sci.* **2**, 313–321 (1998).

Daedalus

Coal, air, fire and water

Coal-mining is a costly, labour-intensive activity. Many attempts have been made to extract at least some of the energy of coal *in situ*, by underground gasification. Daedalus is now updating this idea.

A coal mine, he points out, fills with water unless constantly pumped out. In a flooded mine the water is under hydrostatic pressure. At the bottom of a mine more than about 2.2 km deep, it would reach supercritical pressure. Heated, it could not boil, but would go to supercritical water — a powerful solvent and reaction medium. If it contained dissolved oxygen, organic materials including coal would burn spontaneously in it. So, says Daedalus, drill two holes down to a sufficiently deep coal seam, lower some sort of robot borer to drill a tunnel connecting the two holes, and fill the whole thing with water. Ignite the coal electrically or pyrotechnically, and pump aerated water down one shaft and up the other. The water flow will then deliver the products to the surface.

What will those products be? In the presence of excess fuel — such as a vast coal seam — supercritical combustion is a very partial business. Toluene, for example, burns in limited oxygen to benzoic acid. Coal will go to big aromatic or quinonoid molecules, with enough oxygen in them to dissolve in the water. The upcoming stream of hot water will boil as the hydrostatic pressure lessens. The steam will contain the more volatile organics, easily separable by fractionation, ideally by stepwise expansion in a steam engine. Bigger molecules will stay in the water; some will crystallize out as it cools. So the mine will yield steam power, plus a stream of useful fuels and chemical feedstocks. Sulphur dioxide in the water will be processed to useful sulphuric acid, and solid ash will mainly stay underground.

A supercritical mine will access a vast amount of coal. The initial narrow tunnel will soon widen laterally as it eats the seam away from between the rocky strata above and below it. As it widens, the tunnel will slowly collapse in the middle by subsidence, giving two tunnels. These will gradually move apart through the seam, scouring it out as subsidence propels them sideways. The vast coal reserves more than 2.2 km down, too deep for normal mining, will be accessible at last.

David Jones

The Further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

A pygostyle from a non-avian theropod

The independent evolution of a bird-like tail has been discovered in an oviraptorosaur.

In birds, vertebrae at the tail end of the backbone are fused into a structure called a pygostyle. We have found a pygostyle from a non-avian theropod, an oviraptorosaur from the Upper Cretaceous Nemegt Formation of Bugin Tsav, Mongolia. Although this taxon has similarities to birds, it is likely that the pygostyle evolved independently.

The specimen (Mongolian Geological Institute GIN 940824) includes most vertebrae, along with chevrons, ribs, gastralia, pelvis (Fig. 1), femora, tibiae, fibulae and proximal tarsals. It has a shorter tail (24 segments) than known oviraptorosaurs, and the last five caudals are co-ossified into a pygostyle.

GIN 940824 can be identified as an oviraptorosaur¹ on account of its pneumatic proximal caudal centra, short distal caudal centra, ilia that are close dorsally, and its vertical pubis. It is one of the largest oviraptorids known. Unlike *Ingenia*, recovered from the same site, the preacetabular blade of the ilium is higher than the postacetabular region, as in *Chirostenotes*². It has a shorter tail than *Oviraptor* (GIN 100/42) and a lower preacetabular iliac blade with less pronounced dorsal curvature. The shaft of the pubis is straighter than in *Ingenia*. The ventrally directed pubis is 171% of the ischial length, compared with 148% in *Oviraptor*, 121% in *Ingenia*, and 221% in *Chirostenotes*³. The ischium curves strongly posteriorly, as in *Chirostenotes*^{2,3}. Similarities of the ilium and ischium with North American caenagnathids, a family also known in Asia⁴, suggest it may be referable to that family.

Most non-avian theropods have long tails with elongate (relative length to width) caudal centra, whereas oviraptorosaurs have short tails with short, broad vertebral centra. The minimum caudal counts are 32 for *Conchoraptor* (GIN 110/19), 27 for *Ingenia* and 27 for "*Oviraptor*" *mongoliensis*. GIN 940824 has only 24 caudals, fewer than any non-avian theropod except *Caudipteryx*⁵, which has 22.

The pygostyle is a relatively straight, tapering co-ossified mass. The morphology of caudals 15 to 24 differs from their counterparts in other oviraptorosaurs. Fusion between the twentieth and twenty-first caudals is restricted to dorsal parts of the centra, leaving ventral margins separate. As in some modern birds⁶, the neural arches of the first centra in the pygostyle are not co-ossified. The remaining intervertebral articulations are completely co-ossified, although lines of fusion are still evident. At least two haemal

arches are fused. The number of fused vertebrae is the same as the number of caudals supporting tail feathers (rectrices) in *Caudipteryx*⁵ and modern birds.

We can reject the possibility of pathology because the caudal vertebrae undergo a steady reduction in dimensions from the base to the end of the tail. None of the bones shows rough surface texture, asymmetry, growth interruption or any indication of trauma.

Birds generally have 18 (refs 7,8) to 23

(ref. 9) caudal vertebrae, although most are fused into the synsacrum and pygostyle, leaving between five and nine free caudals¹⁰. The avian pygostyle is composed of centra, neural arches and haemal arches⁹. Although most avian pygostyles have dorsally curved axes, some birds, including flightless ratites and penguins, have straight pygostyles like that of GIN 940824.

Phylogenetic analysis does not place oviraptorosaurs particularly close to bird origins, leaving no reason to assume that the short tail and fused caudals of GIN 940824 are homologous with avian tails. However, feathers are widespread in theropods⁵, possibly including oviraptorosaurs¹¹, so GIN 940824 may have had rectrices similar to those of *Protarchaeopteryx* and *Caudipteryx*⁵. The latter is probably a basal oviraptorosaur, as it has an edentulous, beak-like dentary that is deeply invaded posteriorly by the enlarged external mandibular fenestra, a posteriorly concave ischium, short caudal centra, and a short tail.

Although the terminal vertebrae of *Caudipteryx* are not fused, they seem to form a stiffened rod. The short tail and pygostyle of GIN 940824 maybe indicative of a fan of elongate tail feathers (rectrices) in oviraptorosaurs (Fig. 2) that is more derived than in *Caudipteryx*. Pygostyle-like structures could have evolved independently at least three times in theropods, although the presence of rectrices in two of these taxa suggests a functional association.

Rinchen Barsbold*, **Philip J. Currie†**, **Nathan P. Myhrvold‡**, **Halszka Osmólska§**, **Khishigjav Tsogtbaatar***, **Mahito Watabe¶**

*Geology Institute,

Academy of Sciences of Mongolia,

Ulan Bator 210351, Mongolia

†Royal Tyrrell Museum of Palaeontology, Box 7500, Drumheller, Alberta T0J 0Y0, Canada

‡Microsoft Research, One Microsoft Way, Redmond, Washington 98052-6399, USA

§Polska Akademia Nauk, Instytut Paleobiologii, Ul. Twarda 51/55, 00-818 Warszawa, Poland

¶Hayashibara Museum of Natural Sciences, Shimoishii 1-2-3, Okayama 700-0907, Japan

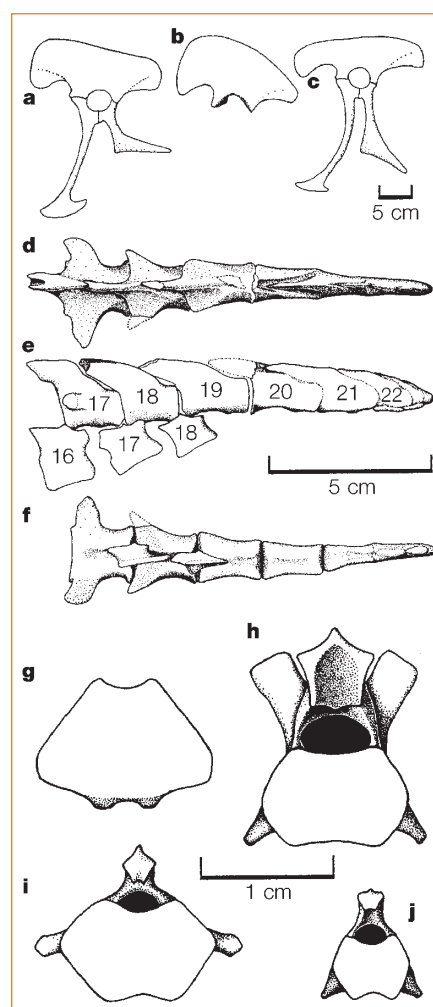


Figure 1 Pelvic girdles of oviraptorids. **a**, GIN 940824; **b**, ilium of "*Oviraptor*" *mongoliensis*, GIN 100/32A; **c**, ilium, pubis and ischium of *Ingenia*, GIN 100/33. **d-f**, Drawings of pygostyle of GIN 940824 in dorsal (**d**), left lateral (**e**) and ventral (**f**) aspects. **g-j**, Posterior views of oviraptorid caudal vertebrae: **g**, nineteenth caudal of GIN 940824; **h**, nineteenth caudal of *Ingenia*, GIN 100/33; **i**, twenty-seventh caudal of *Ingenia*, GIN (100/32); **j**, twenty-sixth caudal of "*Oviraptor*" *mongoliensis*, GIN 100/32A.

Figure 2 (above right) The function of the pygostyle of the new oviraptorosaur interpreted as a support for a fan of tail feathers. Drawing by Michael Skrepnick.



1. Barsbold, R. in *The Encyclopedia of Dinosaurs* (eds Currie, P. J. & Padian, K.) 505-509 (Academic, San Diego, 1997).

2. Currie, P. J. & Russell, D. A. *Can. J. Earth Sci.* **25**, 972-986 (1998).

3. Sues, H.-D. *J. Vert. Paleontol.* **17**, 698-716 (1997).

4. Currie, P. J. et al. *Can. J. Earth Sci.* **30**, 2255–2272 (1993).
5. Ji, Q. et al. *Nature* **393**, 753–761 (1998).
6. Peck, J. L. *J. Morphol.* **3**, 127–136 (1889).
7. Steiner, H. *Vierteljahr. Naturf. Gesell. Zürich* **83**, 279–300 (1938).
8. Baumel, J. J. *Adv. Anat. Embryol. Cell Biol.* **110**, 1–115 (1988).

9. Parker, W. K. *Proc. R. Soc. Lond.* **43**, 465–482 (1888).
10. Oort, van, E. D. *Tijdschr. Ned. Dierk. Vereen.* **9**, 1–144 (1904).
11. Hopp, T. & Orsen, M. in *Dinofest International Symposium, Program and Abstracts* (eds Wolberg, D. L. et al.) 27 (Academy of Natural Sciences, Philadelphia, 1998).

Evolutionary fitness

Tall men have more reproductive success

Sexual selection is a well established evolutionary process based on preferences for specific traits in one sex by members of the other sex. It is important in the evolution of morphological traits, and several sexually dimorphic traits in humans, such as facial hair and facial shape¹, are assumed to be the outcome of such a process. Here we demonstrate that taller men are reproductively more successful than shorter men, indicating that there is active selection for stature in male partners by women.

There are well known academic², social^{3,4}, health^{5–9} and economic^{10,11} correlates of height, but we know of no studies that have examined its direct implications for evolutionary fitness. To examine the possible evolutionary consequences of stature, we have analysed data from the medical records for 4,419 healthy men aged 25–60 who received compulsory medical examinations between 1983 and 1989 at the Lower Silesian Medical Centre in Wrocław, Poland. Because the records were not anonymous and many bachelors admitted to having offspring, the risk of false declarations was probably small.

To avoid any confounding pathological effects, we discarded the data from men whose height was more than three standard deviations from the sample mean (172 ± 6.6 cm). Locality of residence had a significant effect on stature (analysis of variance (ANOVA): $F_{3,4409} = 11.02$, $P < 0.001$, with city-dwellers being taller than rural men), so we selected only individuals resident in either rural villages or in cities with more than 100,000 inhabitants (the extreme sub-populations). This yielded a final sample size of 3,201.

Stature is also confounded by a secular trend (linear regression against age: $b = -0.00157$, $F_{1,3198} = 184.5$, $P < 0.0001$), so we calculated a residual from the common regression line against age for each subject and standardized this against the overall mean for the subject's residence. Stature is also confounded by educational achievement, probably because this correlates with family wealth and status, so we controlled for education.

Figure 1a shows means and variances in stature residuals for men with and without children. When all other variables are held constant, childless men are significantly

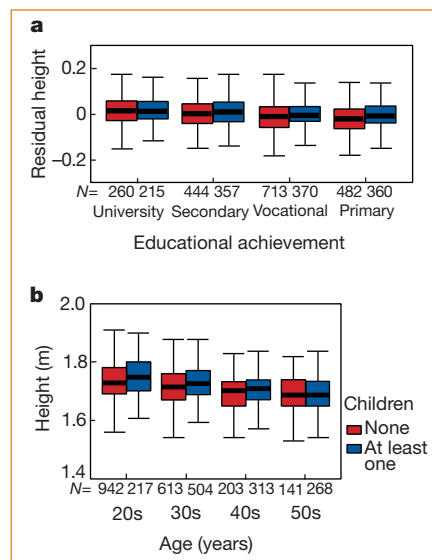


Figure 1 Height (mean $\pm$ s.d. and 95% range) of men with and without children. **a**, Residual height for men with different educational experience (removing effects due to both location of residence and a secular trend over time). **b**, Absolute height for men of different age cohorts, as a function of whether or not they had any children.

shorter than those who have at least one child (ANOVA: childlessness, $F_{1,3198} = 25.5$, $P < 0.001$; education, $F_{1,3198} = 93.1$, $P < 0.001$). Multiple regression with number of children as the dependent variable, with height and age as independent variables, provides quantitative confirmation of these results (one-tailed tests: city, $r^2 = 0.136$, $n = 1,826$; height, $P < 0.001$; age, $P < 0.001$; rural, $r^2 = 0.208$, $n = 1,297$; height, $P = 0.041$; age, $P < 0.0001$).

Comparisons of means for individual age cohorts (Fig. 1b) reveals that men with children are significantly taller than childless men in each case (twenties, $t_{1157} = -2.97$, $P = 0.005$; thirties, $t_{1115} = -3.49$, $P = 0.001$; forties, $t_{514} = -3.06$, $P = 0.002$), except for men in their fifties ($t_{409} = 0.17$, $P = 0.863$). Because these men were born during the 1930s, they entered the marriage market shortly after the Second World War when the population sex ratio was highly skewed in favour of women and sexual selection on males would have been greatly reduced as a result: the sex ratio for adults of working age (18–64 for men, 18–60 for women) in Wrocław was 114.3 women to 100 men in the post-war decade, but fell to 104–105:100 in subsequent decades¹².

These results indicate that the effect of height on reproductive output might be due to shorter men being disadvantaged in

the search for a mate. This idea is supported by the fact that bachelors were significantly shorter than married men (ANOVA with residual height as the dependent variable, childlessness as the covariate, and marital status as the independent variable: $F_{1,3198} = 7.82$, $P = 0.005$).

B. Pawłowski*, R. I. M. Dunbar†, A. Lipowicz‡

*Department of Anthropology, University of Wrocław, ul. Kuznicza 35, 50-138 Wrocław, Poland

†ESRC Research Centre in Economic Learning and Social Evolution, School of Biological Sciences, Nicholson Building, University of Liverpool, Liverpool L69 3BX, UK

e-mail: rimd@liv.ac.uk

‡Institute of Anthropology, Polish Academy of Sciences, ul. Kuznicza 35, 50-951 Wrocław, Poland

1. Perret, D. et al. *Nature* **394**, 884–887 (1998).
2. Teasdale, T. W., Owen, D. R. & Sørensen, T. I. A. *Human Biol.* **63**, 19–30 (1991).
3. Macintyre, S. & West, P. *Sociol. Health Illness* **13**, 149–166 (1991).
4. Bielicki, T. & Szklarska, A. *J. Biosoc. Sci.* **31**, 525–536 (1999).
5. Holl, R. W., Schroder, H. & Heinze, E. *Deutsch. Med. Wochenschr.* **116**, 928–934 (1991).
6. Barker, D. J. P. et al. *Ann. Hum. Biol.* **17**, 1–6 (1990).
7. Bogin, B. *Patterns of Human Growth* (Cambridge Univ. Press, 1988).
8. Herbert, P. R. et al. *Circulation* **88**, 1437–1443 (1993).
9. Kee, F. et al. *Int. J. Epidemiol.* **26**, 748–756 (1997).
10. Komlos, J. *Stature, Living Standards and Economic Development* (Univ. Chicago, 1994).
11. Schumacher, A. J. *Hum. Evol.* **11**, 697–701 (1982).
12. *Rocznik Statystyczny Miasta Wrocławia* (Miejski Urząd Statystyczny we Wrocławiu, Wrocław, 1971).

Physiology

Eel fat stores are enough to reach the Sargasso

It has long been assumed that the European eel (*Anguilla anguilla*) migrates to the Sargasso Sea — a region of the Atlantic Ocean between the Azores and the West Indies — to spawn^{1–3}. During the past decade, however, the number of glass eels has inexplicably dropped⁴, and it has been suggested that a shortage of fat stores in adults, resulting from diminished food resources for juveniles in inland waters, may prevent the starving silver eels from reaching the spawning grounds^{4–6}. But we find that the energetic cost of the 6,000-km migration is actually quite low, with 60% of the fat store remaining available for the developing gonads.

Silver eels leaving the coasts of Europe between September and November are likely to reach the Sargasso Sea from February to June³, so the average swimming speed for a female silver eel 1 metre long is about half a body-length per second. The estimated energy required is around 30% of the total energy at the start⁷. The fat reserves of migrating silver eels range from

Table 1 Energy cost of swimming of migrating silver eels

Mean O ₂ consumption (ml O ₂ kg ⁻¹ h ⁻¹)	Energy consumption (cal g ⁻¹ h ⁻¹)	Swimming distance (km)	Energy cost of swimming (cal g ⁻¹ km ⁻¹)
46.13 (± 9.90)	0.222 (± 0.048)	387.1 (± 11.27)	0.137 (± 0.026)

Values are mean (± s.d.) for five eels.

10 to 28% (ref. 5), which suggests that most European silver eels cannot finish the journey.

We therefore measured the energy consumption of silver eels (1 m long) that were swimming continuously at half a body-length per second for 10 days (Fig. 1). We converted the oxygen-consumption data to fat oxidation by using the oxy-caloric value of fat⁹, the only suitable fuel. The eels' fat-consumption rates when resting and swimming were 10.11 ± 0.36 and 23.06 ± 0.41 mg fat per kg per hour, respectively.

We can use these values to calculate the energy costs for migrating eels. The fat content of silver eels ranges from 10 to 28% with a mean of 20% (ref. 5). Assuming that a 1-m adult silver eel weighing 2 kg, of which 400 g is fat, migrates 43.2 km a day (at half a body-length per second), it will take 139 days for it to reach the Sargasso Sea. During this time it will use about 154 g fat, corresponding to 38.5% of its fat stores. It has been suggested¹⁰ that the energy cost is between 0.329 and 0.417 cal per g per km, some 2.4–3.0 times higher than our results of 0.137 cal per g per km (Table 1). The energy costs may actually be even lower because silver eels migrate at lower temperatures and may be using westward currents.

According to our data, about 60% of the eel's initial fat reserve can be used for gonad development at the end of the journey. Based on a mean energy content of fish eggs of 23.48 kJ per g dry weight¹¹, a 2-kg female silver eel would be able to produce 413 g of eggs. This corresponds to a gonad-somatotrophic index of 22, which is a normal value for hormone-treated animals⁷.

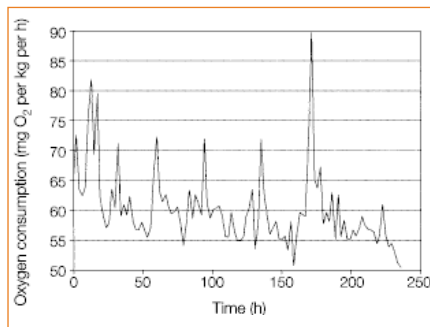


Figure 1 Oxygen consumption of a silver eel (1.5 kg, 90 cm long), swimming at half a body-length per second, measured continuously between 90 and 80% air saturation. Adult migrating silver eel were studied in a 127-litre swim tunnel⁸, with a water temperature of 14 °C and salinity of 33‰. The swim tunnel was calibrated using a laser doppler method at the Hydraulics Laboratory TU, Delft. There was a linear relation between the revolution rate of the propeller and the water velocity, which enabled us to apply a water velocity of half a body-length per second for every eel.

Vincent J. T. van Ginneken,
Guido E. E. J. M. van den Thillart

*Institute of Ecological and Evolutionary Sciences,
Integrative Zoology, van der Klaauw Laboratorium,
PO Box 9511, 2300 RA Leiden, The Netherlands*

- Schmidt, J. *Nature* **111**, 51–54 (1923).
- Miller, M. & McCleave, J. D. *J. Mar. Res.* **52**, 743–772 (1994).
- Fricke, H. & Kaese, R. *Naturwissenschaften* **82**, 32–36 (1995).
- Castonguay, M. *et al. Fish. Oceanogr.* **3**, 197–203 (1994).
- Svedäng, H. & Wickström, H. *J. Fish. Biol.* **50**, 475–486 (1997).
- Tesch, F. W. *The Eel. Biology and Management of Anguilloid Eel* (Chapman & Hall, London, 1977).
- Boëtius, I. & Boëtius, J. *Dana* **1**, 1–28 (1980).
- van Dijk, P. L. M. *et al. J. Fish. Biol.* **42**, 661–671 (1993).
- Elliot, J. M. & Davison, W. *Oecologia* **19**, 195–201 (1975).
- Schmidt-Nielsen, K. *Science* **177**, 222–228 (1972).
- Wootton, R. J. *Symp. Zool. Soc. Lond.* **44**, 133–159 (1979).

Orthography

Children learn an untaught rule of spelling

Children learn the rules of spelling (orthography) through teaching, and also by themselves through reading. The relative influence of these two kinds of learning is an important issue in the study of reading^{1,2}, but it is hard to determine because children have both types of experience with most rules. Here we show that children can learn a sophisticated orthographic rule for themselves, without the help of teaching.

In English, some past-tense verbs ending in 'd' or 't' are spelled with an '-ed' ending, but others are spelled phonetically (for example, 'told' and 'slept'). It takes English children several years to learn when to use '-ed' to represent the past-tense morpheme and when to write them phonetically³. They may learn to do this by rote, but another possibility is that they learn an untaught rule for verb endings.

The rule is that verbs whose stems sound the same in present and past forms (such as 'clear' and 'peel') are given the '-ed' spelling ('cleared' and 'peeled'), and verbs whose stems sound different in the present and past (such as 'hear' and 'sleep') have phonetically spelled endings ('heard' and 'slept'). This stem-based rule is not taught in English schools.

In two experiments, we asked eight- and nine-year-old children (102 children in the first experiment and 90 in the second) to write 'pseudo-verbs' (made-up verbs) which either had the same stem in the present and past or had different stems. If children knew the stem-based rule, they should put '-ed' endings on the past tense

of pseudo-verbs that have the same stem as in the present tense more often than on those that do not.

We gave the children written passages (24 passages in the first experiment, 18 in the second) in which there was a missing word, denoted by a gap. We then read out the whole passage, including the missing word, which the child was asked to write in the gap.

In each passage, the pseudo-verb occurred three times, twice with its present-tense stem, and the third time in the past tense. The missing word, for example 'chaild', which the children had to write, was always the past pseudo-verb. Sometimes the stem sounded the same in the present and past tenses: 'Harry is a chailer. At the moment he is chailing the teacher's book. He ___ another one this morning.' But sometimes the present and past stems sounded different: 'Harry is a cheller. At the moment he is chelling the teacher's book. He ___ another one this morning.' If children follow the stem-based rule, they should spell the missing word as 'chailed' in the first passage and as 'chaild' in the second passage.

The children did use the '-ed' ending to spell the past-tense morpheme more often in passages with the same stem than in those with different stems (71.92% versus 50.83 in the first experiment, and 64.67% versus 54% in the second). They also spelled the ending phonetically more often in the different-stem than in the same-stem passages (36.67% versus 17.17% in the first experiment, 34.50% versus 21.37% in the second).

Our discovery that many eight- and nine-year-old children use a morphemically based, but entirely untaught, orthographic rule has important implications for theories of spelling. The dominant theoretical approach to the study of spelling has been the dual-route model, which claims that people spell words either by converting sounds into letters, or through a lexical route in which spellings of known words are retrieved whole from memory⁴.

The idea of whole-word retrieval has been criticized^{5,6} on the basis that the lexical route involves analysis of words into morphemes. Our data support this morphemic view, and indicate that morphemic structure has a radical effect on the way that children spell known and unknown words.

Peter Bryant*, Terezinha Nunes†, Rowena Snaith‡

*Department of Experimental Psychology,
University of Oxford, Oxford OX1 3UD, UK
e-mail: peter.bryant@psy.ox.ac.uk

†Department of Child Development and Learning,
Institute of Education, University of London,
20 Bedford Way, London WC1H 0AL, UK

‡Department of Psychology, University of Surrey,
Guildford GU2 5XH, UK

1. Frith, U. in *Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading* (eds Patterson, K. E., Marshall, J. C. & Coltheart, M.) 301–330 (Erlbaum, Hove, 1985).
2. Share, D. L. *Cognition* **55**, 151–218 (1996).
3. Nunes, T. et al. *Dev. Psychol.* **33**, 637–649 (1997).
4. Castles, A. & Coltheart, M. *Cognition* **46**, 159–180 (1993).
5. Taft, M. *Reading and the Mental Lexicon* (Erlbaum, Hove, 1991).
6. Caramazza, A. *Issues in Reading, Writing and Speaking: A Neuropsychological Approach* (Dordrecht, Kluwer, 1991).

Environmental genetics

Rapid chromosomal evolution in island mice

Madeira is a small volcanic island in the Atlantic Ocean with steep mountains separating narrow valleys that are the only areas habitable by humans and their commensals. Here we show that house mice (*Mus musculus domesticus*) on Madeira have an unexpected chromosomal diversity, the evolution of which is independent of adaptive processes, relying instead on geographic isolation and genetic drift.

The evolution of animal species living on islands has given rise to spectacular examples of adaptive radiation in response to new physical and biotic environments^{1,2}. Natural selection is the main mechanism responsible for shaping these radiations, leading to accelerated rates of speciation, but non-adaptive processes may also play a part^{1,3}. We suggest that the extreme topography of Madeira has restricted the movement of mouse populations between valleys, setting the stage for extensive chromosomal radiation.

We found six distinct chromosomal races of the house mouse on Madeira, separated

from each other by mountain barriers (Fig. 1). The races have severely reduced diploid numbers relative to the standard for the species ($2N=40$) owing to the presence of numerous robertsonian chromosomal fusions⁴, each of which results from the joining of a pair of the standard chromosomes. These fusions have been recorded in wild house mice⁴, but the six races on Madeira that carry robertsonian fusions differ from each other and from races elsewhere as a result of the particular combinations of fusions that characterize them (Fig. 2).

House mice are thought to have been introduced onto Madeira following the first Portuguese settlement during the fifteenth century, although fourteenth-century charts indicate that the island was known about before then⁵. The extent of their chromosomal differentiation indicates that populations of house mice were isolated in different valleys for long enough to allow fusions to accumulate. Because each successive fusion as it appears is associated with only a slight underdominance⁴, its fixation can most parsimoniously be explained by population processes such as genetic drift, as inferred for other chromosomal races of house mice⁶.

This example of chromosomal variation among mice on Madeira suggests that human movements were similarly restricted, limiting the passive migration of commensal house mice between valleys. So far, not a single hybrid has been recorded between races carrying robertsonian fusions among the 143 mice that we have karyotyped. We found extensive chromosomal heterozygosity in Funchal, however, probably because of the recent immigration

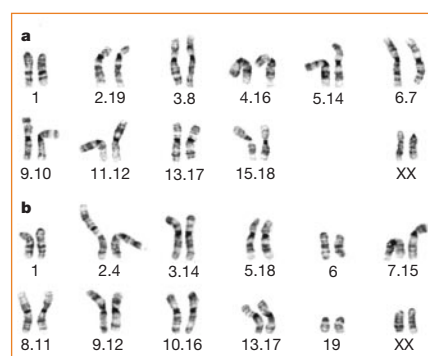


Figure 2 Karyotypes of female mice from the two most differentiated chromosomal races. **a**, $2N=22$ (race identified by red dots in Fig. 1); **b**, $2N=24$ (green dots in Fig. 1).

of standard diploid mice through the commercial port.

Hybrids between any of the Madeiran mouse races carrying robertsonian fusions would be sterile or infertile owing to the complex chromosomal configurations that would be produced at meiosis^{7,8}. Our results indicate not only that accelerated rates of radiation can occur without involving adaptive processes, but also that chromosomal evolution can be an efficient mechanism of isolation, as several reproductively isolated chromosomal races have appeared in less than 500 years.

Janice Britton-Davidian*, **Josette Catalan***, **Maria da Graça Ramalhinho†**, **Gula Ganem***, **Jean-Christophe Auffray***, **Ruben Capela‡**, **Manuel Biscoito§**, **Jeremy B. Searle¶**, **Maria da Luz Mathias#**

*Laboratoire de Génétique & Environnement, ISEM, UMR 5554 CNRS, Université Montpellier II, 34095 Montpellier cedex 5, France

e-mail: britton@isem.univ-montp2.fr

†Centro de Biologia Ambiental, Museu Nacional de Historia Natural, Rua da Escola Politecnica, 1200 Lisboa, Portugal

‡Universidade de Madeira, Departamento de Biologia, Praça do Município, 9000 Funchal, Madeira, Portugal

§CITMA, Rua da Alameda 78, 9000 Funchal, Madeira, Portugal

¶Department of Biology, University of York, PO Box 373, York YO10 5YW, UK

#Centro de Biologia Ambiental, Departamento de Zoologia e Antropologia, FCUL, Campo Grande Bloco C2, 1700 Lisboa, Portugal

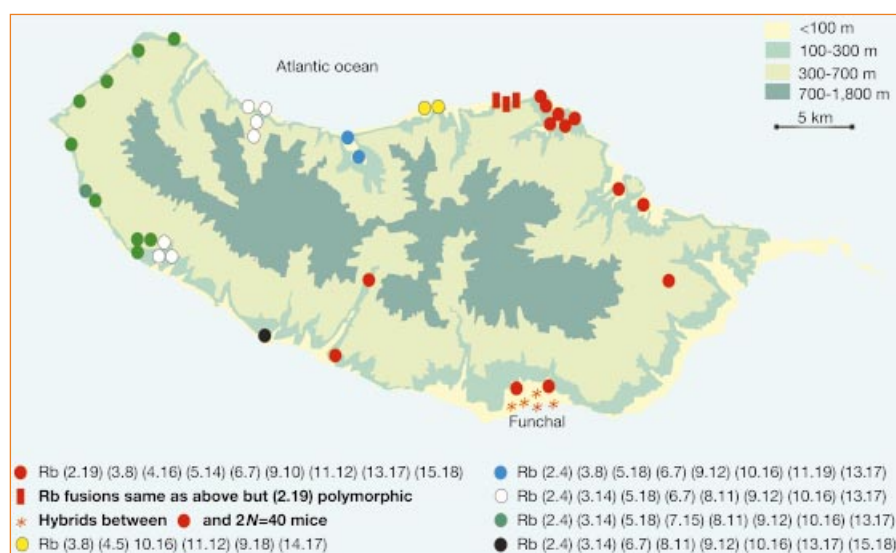


Figure 1 Distribution and composition of chromosomal races in Madeira (32.75° N, 16.97° W). Of the 20 robertsonian (Rb) fusions observed, 13 have not previously been recorded. The presence of two fusions involving chromosome 19 is exceptional: they are the first observed in 91 fusions in wild house mice^{4,9}. Diploid numbers (N) and sample sizes (n) with ranges are as follows: red dot, $2N=22$, $n=43$ (1–21); red rectangle, $2N=23–24$, $n=5$ (1–2); red star, $2N=24–40$, $n=38$ (2–19); yellow dot, $2N=28–30$, $n=5$ (1–4); blue dot, $2N=25–27$, $n=10$ (2–8); white dot, $2N=24–26$, $n=11$ (1–4); green dot, $2N=24–27$, $n=25$ (1–7); black dot, $2N=24$, $n=6$. Map source: Atlas Digital do Ambiente-DGA.

1. Grant, P. R. (ed.) *Evolution on Islands* (Oxford Univ. Press, 1998).
2. Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K. & Rodriguez-Schettino, L. *Science* **279**, 2115–2118 (1998).
3. Provine, W. B. in *Genetics, Speciation, and the Founder Principle* (eds Giddings, L. V., Kaneshiro, K. Y. & Anderson, W. W.) 43–76 (Oxford Univ. Press, New York, 1989).
4. Nachman, M. W. & Searle, J. B. *Trends Ecol. Evol.* **10**, 397–402 (1995).
5. Mathias, M. L. & Mira, A. *Biol. J. Linn. Soc.* **46**, 13–24 (1992).
6. Hauffe, H. C. & Searle, J. B. *Evolution* **47**, 1374–1395 (1993).
7. Baker, R. J. & Bickham, J. W. *Proc. Natl Acad. Sci. USA* **83**, 8245–8248 (1986).
8. Hauffe, H. C. & Searle, J. B. *Genetics* **150**, 1143–1154 (1998).
9. Garagna, S., Zuccotti, M., Redi, C. A. & Capanna, E. *Nature* **390**, 242 (1997).

MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*

Shawn Ahmed & Jonathan Hodgkin

MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK

The germ line is an immortal cell lineage that is passed indefinitely from one generation to the next. To identify the genes that are required for germline immortality, we isolated *Caenorhabditis elegans* mutants with mortal germ lines—worms that can reproduce for several healthy generations but eventually become sterile. One of these *mortal germline* (*mrt*) mutants, *mrt-2*, exhibits progressive telomere shortening and accumulates end-to-end chromosome fusions in later generations, indicating that the MRT-2 protein is required for telomere replication. In addition, the germ line of *mrt-2* is hypersensitive to X-rays and to transposon activity. Therefore, *mrt-2* has defects in responding both to damaged DNA and to normal double-strand breaks present at telomeres. *mrt-2* encodes a homologue of a checkpoint gene that is required to sense DNA damage in yeast. These results indicate that telomeres may be identified as a type of DNA damage and then repaired by the telomere-replication enzyme telomerase.

In most higher organisms, the germ line is responsible for perpetuation of the species. Differentiated germ cells fuse together to form a zygote, which then develops into an organism containing a soma and a new germ line, and this cycle repeats until a species becomes extinct. Thus, the germ line has the ability to proliferate indefinitely and can be thought of as an immortal cell lineage¹.

We decided to study the problem of germline immortality in the nematode *C. elegans*. *C. elegans* worms clearly demarcate their germ cells from the very first zygotic cell division². In addition, *C. elegans* populations are normally composed of self-fertilizing hermaphrodites that are homozygous at most or all genetic loci³. Therefore, a single worm can give rise to generations of descendants that are essentially genetically identical and whose germ lines represent a single continuous immortal cell lineage.

We have identified a number of *C. elegans* mutants with mortal germ lines. The *mrt-2* mutant displays the telomere shortening and late-onset chromosome fusion phenotypes that are seen in 'telomerase-negative' mouse and yeast mutants^{4–6}. Telomerase is a reverse transcriptase that adds repeats to the ends of chromosomes⁷. Five genes are known to be required for telomerase activity *in vivo*⁸. Mutations in these genes result in late-onset sterility/senescence, and end-to-end chromosome fusions that can be explained by progressive telomere shortening^{4–6,8–10}. However, the telomerase-defective *C. elegans mrt-2* mutant has additional phenotypes: it displays weak chromosome loss at all times and is also hypersensitive to agents that damage DNA. These results suggest that *mrt-2* may have a defect in its response to DNA damage (a defect in DNA repair, in a DNA damage checkpoint, or in both).

Some yeast mutants with defects in double-strand break DNA repair (*ku*, *sir* and the *mre11/xrs2/rad50* nuclease mutants) have short telomeres^{11–15}. In addition, a number of mutants that have defects in sensing DNA damage (DNA damage checkpoint mutants) also exhibit telomere defects. For example, deficiencies in the human ATM checkpoint protein result in aberrant telomere shortening^{16,17}, and a number of yeast checkpoint mutants have short but stable telomeres^{18–20}. *mrt-2* encodes the *C. elegans* homologue of the *Schizosaccharomyces pombe rad1*⁺ and *Saccharomyces cerevisiae RAD17* checkpoint genes. Although the *S. pombe rad1* mutant has shorter telomeres than wild type¹⁹, it does not exhibit the telomere catastrophe that we find in the higher eukaryote *C. elegans*.

A screen for mortal germline mutants

To investigate how the germ line achieves immortality, a screen was conducted for *C. elegans* mutants with mortal germ lines. Four hundred lines were established from single F₂ progeny of ethylmethanesulphonate (EMS) mutagenized worms and grown clonally for 16 generations at 25 °C. Sixteen independent mortal germline mutants were identified that become effectively sterile (<2 progeny per worm) between generations F₄ and F₁₆ (Fig. 1). All of these *mrt* mutants behaved as expected for single recessive mutations when they were outcrossed (data not shown). The unexpectedly large number of *mrt* mutants that were identified in our small pilot screen suggests a conservative estimate of ~50 genes that are specifically required for germline immortality in *C. elegans* (the forward mutation frequency per gene is, on average, 1 in every 4,000 F₂ progeny from EMS-mutagenized worms)²¹.

Of these 16 *mrt* mutants, 12 are temperature-sensitive and will grow indefinitely at 15 °C or 20 °C, but will become sterile if they are shifted to 25 °C for several generations (data not shown). There are two possible explanations for the large number of temperature-sensitive mortal germline mutants that we uncovered. First, these

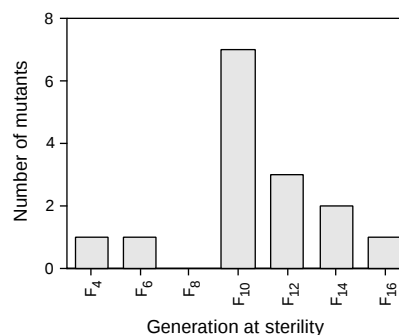


Figure 1 The generation at sterility for 16 mortal germline mutants. Clonal strains were grown on petri dishes at 25 °C and could grow for two generations per dish before starvation. Strains were passaged by transferring six L1 larvae to a fresh dish, which ensured that zygotic or maternal sterile mutations that may have been segregating in the mutant strains would not be mistakenly identified as *mrt* mutations. Lines were scored as 'sterile' when there were less than two progeny per worm per plate.

temperature-sensitive mutations may identify genes that are required to protect the germ line against damage that accumulates only at high temperature. Second, there may be some redundancy in the genes that control germline immortality, such that one only observes a mutant phenotype at high temperature when one branch of a pathway has been knocked out by temperature and the other branch by mutation.

Four of the *mortal germline* mutants that we isolated became sterile at all temperatures (data not shown). These non-temperature-sensitive *mrt* mutants must be outcrossed periodically to

maintain them. One of these mutants, *mrt-2* (e2663), became sterile at generation F₁₄ when it was first identified in our *mortal germline* screen. *mrt-2* was outcrossed several times, and homozygous *mrt-2* lines were re-established. The generation at sterility for 23 different *mrt-2* lines varied widely, from generation F₁₀ to generation F₂₈ (Fig. 2a). However, different F₂ lines from the same F₁ heterozygote all became sterile at about the same time (Fig. 2b; and data not shown). These results suggest that *mrt-2* accumulates some kind of damage that is inherited and segregates evenly amongst the progeny of an outcross. In addition, *mrt-2* produces normal brood sizes at generation F₂ (about 300 progeny per worm), but brood size gradually drops to the point of sterility (Fig. 2b). This gradual drop in brood size indicates that the germ line of *mrt-2* accumulates damage slowly over many generations. *mrt-2* has additional phenotypes, described below, which led us to focus attention on it.

Genome instability in the *mrt-2* mutant

C. elegans populations are normally composed of XX hermaphrodites, but XO males occasionally arise as a result of X-chromosome loss²². *mrt-2* displays a weak high incidence of males (Him) phenotype and produces $0.9 \pm 0.6\%$ males ($n = 25$, F₃ broods) as opposed to the 0.2% normally observed in wild type. This increase in X-chromosome loss suggests a defect in genome stability.

Although the *mrt-2* germ line is not significantly hypersensitive to ultraviolet light (data not shown), it is hypersensitive to X-rays (Fig. 3a) which can damage DNA by inducing double-strand breaks. In another test of the ability of *mrt-2* to withstand double-strand breaks, we used a *C. elegans* mutator strain, *mut-2*(r459), which has high levels of transposon activity in its germ line that are likely to result in double-strand breaks²³. The germ lines of *mrt-2*;*mut-2*(r459) double mutants gave rise to progeny with two to three times the level of lethality expected for the sum of the single mutants

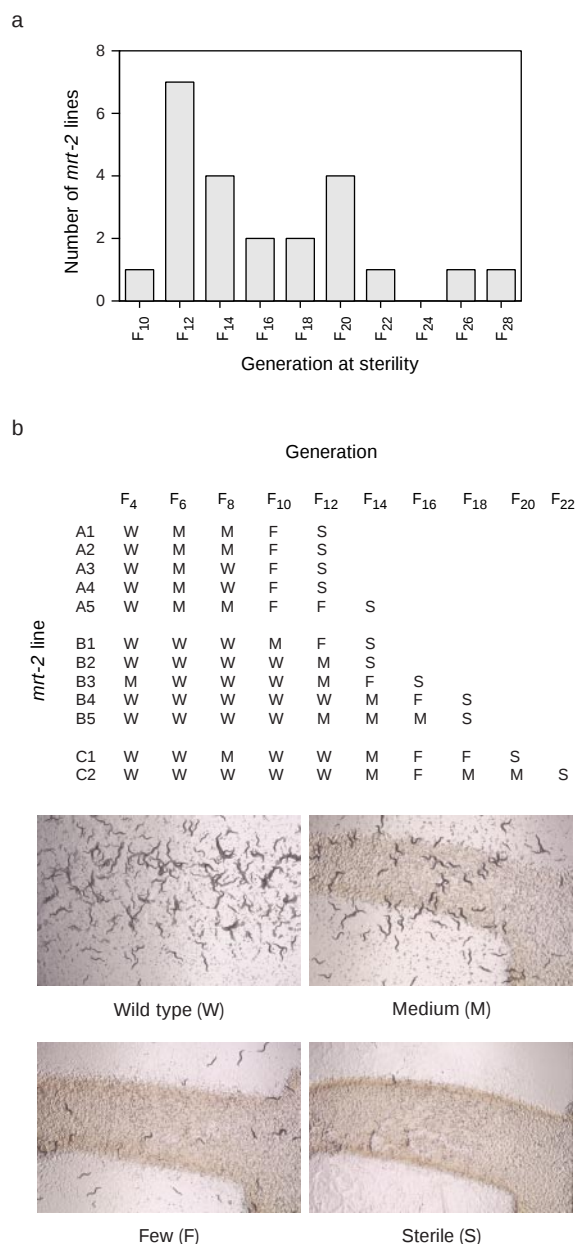


Figure 2 The Mortal Germline phenotype of *mrt-2*. **a**, The generation at sterility for 23 different *mrt-2* F₂ lines. **b**, Brood size drops for 12 *mrt-2* lines. Three different *mrt-2* + F₁ lines (A, B and C) gave rise to the *mrt-2* sibling lines that are shown. Brood size was determined by examining plates seeded with 6 *mrt-2* L1 larvae after 7 days of growth at 20 °C. Plates were scored as wild type (W), medium (M), few (F) and sterile (S). A light-brown *Escherichia coli* lawn used to feed *C. elegans* worms is visible on M, F and S plates. Brood sizes of individual worms at these stages were wild type, 262 ± 115 (4 ± 4 dead eggs); medium, 70 ± 52 (13 ± 12 dead eggs); few, 20 ± 11 (55 ± 31 dead eggs); and sterile, 4 ± 5 (47 ± 43 dead eggs) (10 broods scored at each level of fecundity; 5 each from 2 different lines).

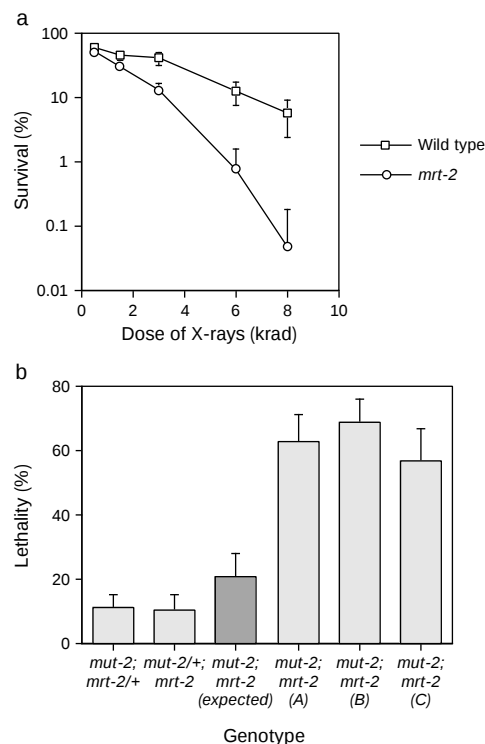


Figure 3 *mrt-2* has defects in responding to DNA damage. **a**, *mrt-2* germ lines are more sensitive to X-rays than are N2 wild-type germ lines (10 broods scored at each dose). **b**, Levels of embryonic lethality in three independent lines (A, B and C) of *mut-2*; *mrt-2* double mutants and their *mut-2*; *mrt-2*+/+ and *mut-2*+/+; *mrt-2* siblings (9–10 broods scored for each column). The expected additive lethality for *mut-2* and *mrt-2* is indicated (dark grey).

(Fig. 3b). The simplest explanation for this synergistic *mrt-2*; *mut-2(r459)* lethality is that *mrt-2* is defective in responding to double-strand breaks caused either by transposition or by X-rays.

Chromosome fusions and telomere shortening

To gain further insight into the genome stability defect of *mrt-2*, *mrt-2* worms were stained with diamidinophenolindole (DAPI) and their chromosomes were examined. Early generation *mrt-2* worms had the six pairs of chromosomes expected for wild-type, whereas late generation *mrt-2* worms contained only three, four or five DAPI-staining chromosome pairs per worm, suggesting the occurrence of chromosome fusions (Fig. 4a). In addition, the number of visible chromosome pairs decreased over the course of eight generations in late-generation *mrt-2* lines (Fig. 4b).

When late-generation *mrt-2* worms were outcrossed, we frequently observed a dominant Him phenotype. This unusual phenotype is known to occur in *C. elegans* as a result of X-autosome

chromosome fusions²⁴. Although in most organisms dicentric chromosome fusions are frequently torn apart during mitosis²⁵, *C. elegans* has holocentric chromosomes²⁶, therefore chromosome fusions are stable and can be mapped genetically²⁴. The dominant Him phenotype from 12 independent *mrt-2* lines consistently mapped to one end of the X chromosome (data not shown). In three out of three cases examined, this phenotype also mapped to the end of an autosome (different in each case) (Fig. 4c), confirming the presence of X-autosome chromosome fusions in late-generation *mrt-2* worms. All chromosome fusions tested were homozygous viable ($n = 10$; data not shown), indicating that they were not missing any essential genes and were therefore probably end-to-end chromosome fusions.

The late-onset end-to-end chromosome fusion phenotype of *mrt-2* suggested a defect in telomere replication^{4–6}. *C. elegans* has telomeres consisting of simple tandem TTAGGC repeats similar to those found in most other eukaryotes²⁷. In Southern blots, the telomeres of wild-type worms appear as a smear that runs from 2 to 7 kilobases (kb) and either stays fairly constant in length at 20 °C (Fig. 5a), or increases in length if worms have short telomeres to begin with (Fig. 5b). In contrast, the telomeres of *mrt-2* worms shorten at a rate of about 12 base pairs (bp) per cell division (about 125 bp per generation and an estimated 10 cell divisions per generation) (Fig. 5). *mrt-2* telomeres shorten progressively at least until generation F₁₈ (Fig. 5c), by which time most *mrt-2* lines have become sterile (Fig. 2a). Occasional telomere signals disappear as *mrt-2* worms are passaged (Fig. 5a, arrow), which may be the consequence of a telomere fusion event, a recombination event, or perhaps the segregation of long and short alleles of a particular telomere.

In initial experiments, *mrt-2* lines always produced X-autosome chromosome fusions that had fused at the right end of the X chromosome ($n = 5$). This result indicates that the right telomere of the X chromosome may be prone to fusion events, perhaps because it is shorter than the left. To test this possibility, we identified several *mrt-2* lines that were homozygous for a short telomere at the left end of the X chromosome. Four independent X-autosome chromosome fusions were recovered from these lines, of which two had fused at the right end and two had fused at the left end of the X chromosome (Fig. 2d, *eT5*; and data not shown). Thus, the left end of the X chromosome will undergo late-generation chromosome fusion events if its initial length is short. These results suggest that telomere shortening causes the late on-set chromosome fusions observed in *mrt-2* (Fig. 2).

A dominant-negative allele of the telomere-binding protein TRF2 can induce end-to-end chromosome fusions in mammalian cells²⁸. This phenotype occurs within a few cell divisions, however, and the resulting telomere–telomere fusions produce an increase in telomere length as observed by Southern blot²⁸. End-to-end chromosome fusions have also been observed in mice mutant for poly(ADP-ribose) polymerase, and telomere length in these mice is short but stable²⁹. These chromosome-fusion phenotypes are distinct from the late-onset chromosome fusion and progressive telomere shortening phenotypes observed in *mrt-2*. Instead, *mrt-2* is similar to the telomerase-negative mouse and yeast mutants, which also experience progressive telomere shortening, late-onset chromosome fusions and late-onset sterility/senescence^{4–6,9}. Although yeast telomerase-negative mutants can use recombination to periodically recover from senescence^{5,30}, we have never observed a *mrt-2* line to recover from its Mortal Germline phenotype (Fig. 2). The late-onset sterility of *mrt-2* lines probably results from increasing numbers of chromosome fusions, which will lead to meiotic non-disjunction and massive aneuploidy.

mrt-2 is a checkpoint gene

The *mrt-2* mutant is defective in responding both to normal double-strand breaks present at telomeres and to abnormal DNA damage

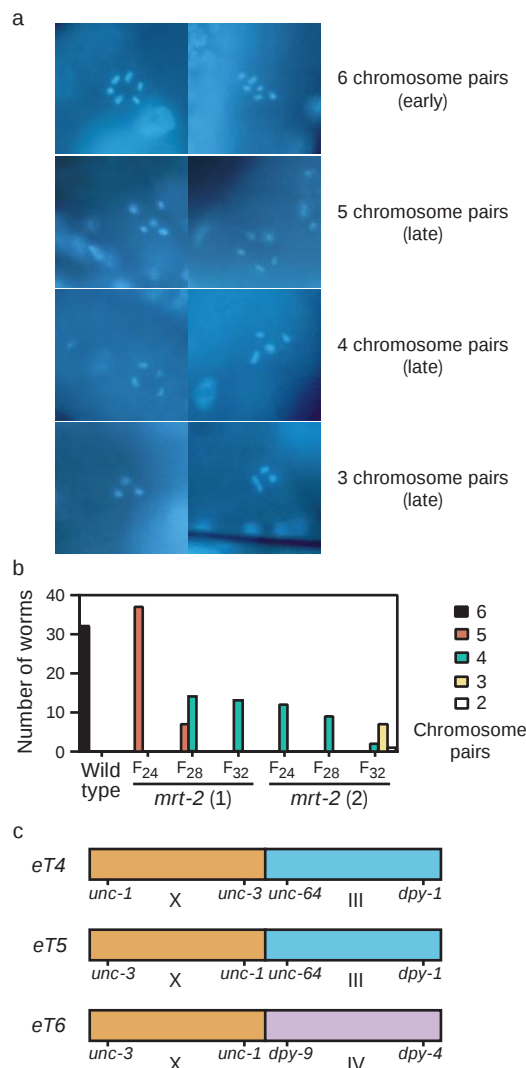


Figure 4 *mrt-2* exhibits late-onset end-to-end chromosome fusions. **a**, DAPI-stained early and late generation *mrt-2* oocyte nuclei arrested at metaphase I in diakinesis (when homologous chromosomes are synapsed together). To ensure that chromosome numbers were accurately determined, three or more oocytes with clearly separated bivalents were scored per worm. These are photographs of representative nuclei in which all chromosome pairs happen to lie in the same plane of focus. **b**, Number of worms containing a particular number of visible chromosome pairs. Wild-type control and successive generations of two late generation *mrt-2* strains (1 and 2) are shown. **c**, X-autosome chromosome fusions isolated from *mrt-2* display pseudo-linkage between an end of the X chromosome and an end of an autosome.

caused by X-rays and transposition. We mapped *mrt-2* between *dpy-18* and *nob-1* on the right arm of chromosome III (Fig. 6a). Blast analysis of this region of the genome revealed a gene *Y41C4A.14*, also known as *hpr-1*, a homologue of the *S. pombe rad1⁺* checkpoint gene^{31,32}. The *S. pombe rad1* mutant is hypersensitive to X-rays³³ and has short telomeres¹⁹. This candidate gene was sequenced in *mrt-2*(e2663) and contained a mutation in the 3' splice junction of the second intron (Fig. 6b). This mutation occurs in a highly conserved AG dinucleotide that is found at the 3' ends of most *C. elegans*

introns. Five independent extrachromosomal arrays containing the wild type *Y41C4A.14* gene were able to rescue the X-ray hypersensitivity, telomere shortening and Mortal Germline phenotypes of *mrt-2* (Fig. 6c, d; and data not shown). In contrast, when *green fluorescent protein* was inserted in-frame between exons 1 and 2 of *Y41C4A.14*, the fusion construct failed to rescue *mrt-2* (data not shown), confirming that *mrt-2* is the *Y41C4A.14* checkpoint gene.

mrt-2 is the *C. elegans* homologue of the *S. pombe rad1⁺* and *S. cerevisiae RAD17* checkpoint genes that are conserved from yeast

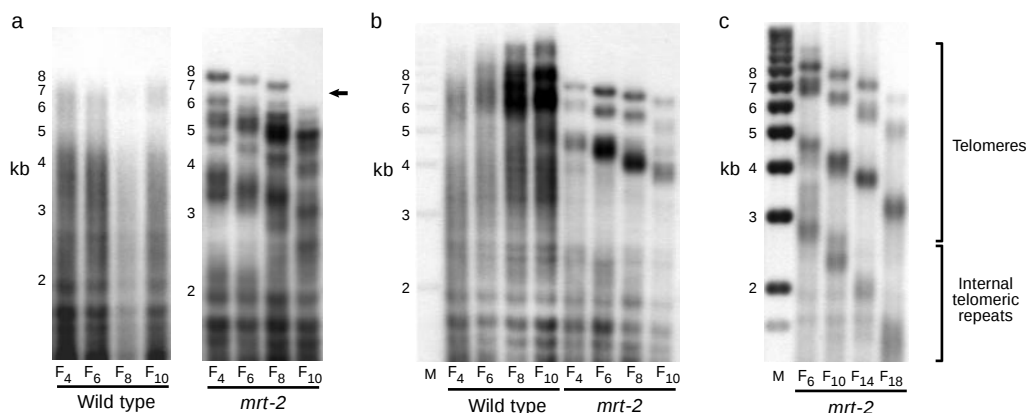


Figure 5 *mrt-2* has a progressive telomere-shortening phenotype. Southern blots were probed with the *C. elegans* telomere sequence (TTAGGC)₂₆. **a**, *mrt-2* telomeres shorten progressively. An arrow marks an *mrt-2* telomere that vanishes in this lineage. **b**, If a *C. elegans* strain with short telomeres, such as *mrt-2*, is crossed with a wild-type strain and

wild-type *F₂* are picked, telomeres from the parent with short telomeres will elongate as they equilibrate back to normal lengths. In contrast, telomeres of a sibling *mrt-2* strain all shorten progressively. **c**, Telomere length of a *mrt-2* strain grown until generation *F₁₈*.

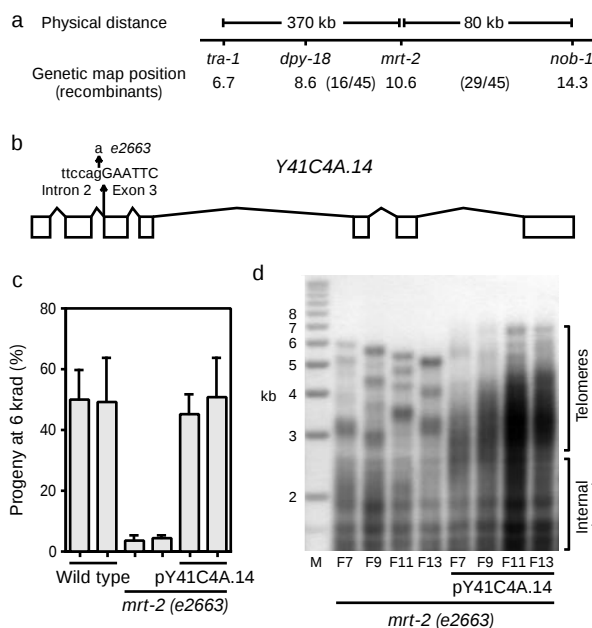
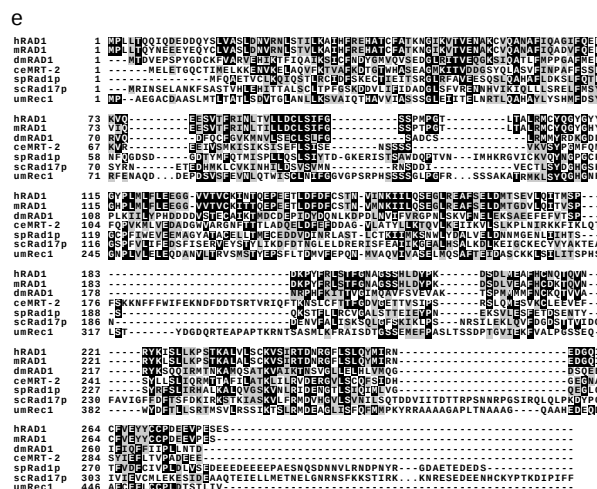


Figure 6 *mrt-2* encodes a homologue of the *S. pombe rad1⁺* and *S. cerevisiae RAD17* checkpoint genes. **a**, *mrt-2* was mapped between *dpy-18* and *nob-1* using three-factor crosses⁵⁰, by picking Dpy-non-Nob and Nob-non-Dpy *F₂* progeny from *dpy-18nob-1/ mrt-2* *F₁* and scoring the X-ray hypersensitivity phenotype in the *F₃* generation. The number of recombinants picked between these genes is indicated in parentheses. Although *mrt-2* maps closer to *dpy-18* than to *nob-1*, *mrt-2* is only 80 kb from *nob-1* indicating that there may be some distortion of genetic distance near *nob-1*. **b**, *mrt-2* (e2663) contains a G to A mutation in the splice acceptor of intron 2 of the *Y41C4A.14/hpr-1* checkpoint gene. **c**, Two extrachromosomal arrays containing a single 7-kb gene—the wild type *Y41C4A.14* gene (pY41C4A.14)—rescue the X-ray hyper-



sensitivity of *mrt-2*(e2663) strains. Non-rescued *mrt-2* sibs and N2 wild type controls are shown (5 broods scored for each column). **d**, Telomere length in a *mrt-2* strain carrying an extrachromosomal array containing pY41C4A.14. A non-rescued *mrt-2*(e2663) sibling strain is shown for comparison. **e**, MRT-2 homologues. Identities are highlighted in black and conservative substitutions in grey. The proteins are human RAD1 (hRAD1), mouse RAD1 (mRAD1), *D. melanogaster* RAD1 (dmRAD1), *C. elegans* MRT-2 (ceMRT-2), *S. pombe* Rad1p (spRad1p), *S. cerevisiae* Rad1p (scRad1p) and *U. maydis* Rec1 (umRec1). Long stretches of non-homologous sequences were omitted (dotted line) for ScRad1p and umRec1. Sequences were aligned using clustalW1.7 and then manually adjusted.

to mammals (Fig. 6e)^{31,32,34,35}. In yeast, these genes are required to delay cell-cycle progression in response to DNA damage or in response to a block in DNA replication^{33,36,37}. As might be expected for a *C. elegans* checkpoint mutant, the germ line of *mrt-2* is defective in responding to DNA damage caused by X-rays and transposition (Fig. 3). In addition, *mrt-2* exhibits the progressive telomere shortening, end-to-end chromosome fusion, and late-onset sterility phenotypes that are seen in 'telomerase-negative' mouse and yeast mutants⁴⁻⁶. Thus, the *mrt-2* checkpoint gene is required for telomere replication. Although several yeast checkpoint mutants including *S. pombe rad1* have telomeres that equilibrate to lengths that are shorter than wild type, they do not exhibit the progressive telomere shortening and senescence phenotypes characteristic of yeast telomerase mutants¹⁹. Curiously, *S. pombe rad1*⁺ is in the same epistasis group as *rad3*⁺ (ref. 33), and a *rad3,tel1* double mutant that is mutant for two related phosphatidylinositol-3-like kinases has a 'telomerase-negative' phenotype *in vivo*⁶. The homologous *S. cerevisiae* double mutant, *mec1,tel1*, also appears to lack telomerase activity *in vivo*²⁰. Although *tel1* mutants do not have major defects in checkpoint function^{38,39}, these results argue that the *rad1*⁺/*RAD17/mrt-2* checkpoint may be required for telomere replication in yeast, and that genetic redundancy may mask this function in single checkpoint mutants.

How might checkpoint proteins recognize telomeres? In *S. cerevisiae*, the *rad1*⁺/*RAD17/mrt-2* checkpoint can delay the cell cycle in response to a single double-strand break⁴⁰. Telomeres are double-strand breaks with short 3' overhangs⁴¹, and these ends fold back and the 3' overhangs are buried in double-stranded telomeric DNA in human cells⁴². These telomeric loops may unfold during S phase to reveal double-strand breaks that could evoke a checkpoint response. Alternatively, a checkpoint response can be triggered in yeast as a result of a block in DNA replication^{36,37}, and it is possible that replication forks stall at telomeres and that these structures are recognized by checkpoint proteins.

Which checkpoint function might be required for telomere replication? Loss of the *S. cerevisiae* telomere-binding protein Cdc13p results in *RAD17*-dependent degradation of the C-rich strand of telomeres⁴³, suggesting that the *RAD17* checkpoint may be involved in the C-strand degradation that is normally observed in late S phase and may be required for telomerase activity⁴⁴. In this context, it is relevant that an allele of *CDC13* is telomerase negative⁴⁵. It is also possible that checkpoint proteins recognize telomeres as a special kind of DNA damage and actively recruit telomerase to initiate repair. In *S. cerevisiae*, the Ku and Sir double-strand-break repair proteins are stored in reservoirs at telomeres but then migrate to double-strand breaks elsewhere in the genome in a checkpoint-dependent manner^{46,47}. In addition, work in the ciliate *Oxytricha* has shown that the telomerase RNA is sequestered away from telomeres during most of the cell cycle, but colocalizes with replicating DNA during S phase⁴⁸.

Our study shows that the problems of telomere maintenance and germline immortality can be addressed using an unbiased forward genetic approach in *C. elegans*. Our main conclusion is that a deficiency for a single DNA damage checkpoint protein can result in a telomerase-negative phenotype *in vivo*. This finding was unanticipated from studies of yeast checkpoint mutants¹⁹ and has implications for genome stability in germline and somatic cells of other higher eukaryotes such as mammals. For example, mice that are deficient in both telomerase and the p53 checkpoint protein display high levels of cancer⁴⁹; thus, the genome instability resulting from a telomerase deficiency has serious consequences in the absence of a response to DNA damage. The checkpoint protein described here, MRT-2, is required both for telomere replication and for response to DNA damage. If the human MRT-2 homologue, RAD1, shares these characteristics, then its loss would result in a double-edged sword: in this context, we note the presence of a tumour suppressor gene near the human RAD1 locus^{31,32,34}. □

Methods

Strains

C. elegans strains were mutagenized, cultured and crossed as described⁵⁰. All experiments were carried out at 20 °C unless otherwise stated.

X-ray hypersensitivity

Young L4 larvae were irradiated in a Torrex X-ray machine at 143kV, and larvae were picked to separate plates and transferred after 48 h and again after 24 h. Plates were scored for adult progeny 36 h after transfer. The number of progeny from irradiated worms was compared with that of unirradiated siblings to give the percentage of survival after irradiation, providing a measure of how sensitive a worm's germ line is to X-rays.

mrt-2 molecular genetics

The *Y41C4A.14* gene was amplified by polymerase chain reaction (PCR) from multiple wild-type and *mrt-2(e2663)* strains and sequenced directly to ascertain the presence of the *mrt-2(e2663)* splice-junction mutation. The wild-type *Y41C4A.14* gene including its 4-kb promoter was PCR-amplified from a *Y41C4* yeast artificial chromosome (YAC) miniprep for subcloning. Plasmid DNA was injected at 0.5 ng μL^{-1} pY41C4A.14 (*FspI*-linearized), 0.5 ng μL^{-1} pCes1943 (*rol-6*, gift of Diana Janke) (*NruI*-linearized) and 50 ng μL^{-1} N2 genomic DNA (*PvuII*-linearized).

Telomere length

Genomic DNA was prepared using a Puregene DNA isolation kit (Gentra). *HinFI*-digested genomic DNA was separated on 0.6% agarose gels at 1.5V cm^{-1} , and Southern blotting was carried out using a digoxigenin-dUTP-labelled PCR probe according to manufacturer's protocols (Boehringer Mannheim). The probe was made using T3 and Te12 (5'-GAATAATGAGAATTTTCAGGC-3') primers to amplify telomeric repeats from the cTel5X plasmid²⁷ using PCR.

Received 17 August; accepted 29 October 1999.

- Wylie, C. Germ cells. *Cell* **96**, 165–174 (1999).
- Strome, S. & Wood, W. B. Immunofluorescence visualization of germ-line-specific cytoplasmic granules in embryos, larvae, and adults of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **79**, 1558–1562 (1982).
- Hodgkin, J. & Doniach, T. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics* **146**, 149–164 (1997).
- Blasco, M. A. *et al.* Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
- Nakamura, T. M., Cooper, J. P. & Cech, T. R. Two modes of survival of fission yeast without telomerase. *Science* **282**, 493–496 (1998).
- Naito, T., Matsuura, A. & Ishikawa, F. Circular chromosome formation in a fission yeast mutant defective in two ATM homologues. *Nature Genet.* **20**, 203–206 (1998).
- Greider, C. W. & Blackburn, E. H. The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**, 887–898 (1987).
- Nugent, C. I. & Lundblad, V. The telomerase reverse transcriptase: components and regulation. *Genes Dev.* **12**, 1073–1085 (1998).
- Lundblad, V. & Szostak, J. W. A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**, 633–643 (1989).
- Lee, H. W. *et al.* Essential role of mouse telomerase in highly proliferative organs. *Nature* **392**, 569–574 (1998).
- Palladino, F. SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell* **75**, 543–555 (1993).
- Porter, S. E., Greenwell, P. W., Ritchie, K. B. & Petes, T. D. The DNA-binding protein Hdf1p (a putative Ku homologue) is required for maintaining normal telomere length in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **24**, 582–585 (1996).
- Boulton, S. J. & Jackson, S. P. Identification of a *Saccharomyces cerevisiae* Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res.* **24**, 4639–4648 (1996).
- Boulton, S. J. & Jackson, S. P. Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J.* **17**, 1819–1828 (1998).
- Nugent, C. I. *et al.* Telomere maintenance is dependent on activities required for end repair of double-strand breaks. *Curr. Biol.* **8**, 657–660 (1998).
- Vaziri, H. *et al.* ATM-dependent telomere loss in aging human diploid fibroblasts and DNA damage lead to the post-translational activation of p53 protein involving poly(ADP-ribose) polymerase. *EMBO J.* **16**, 6018–6033 (1997).
- Sprung, C. N., Bryan, T. M., Reddel, R. R. & Murnane, J. P. Normal telomere maintenance in immortal ataxia telangiectasia cell lines. *Mutat. Res.* **379**, 177–184 (1997).
- Greenwell, P. W. *et al.* TEL1, a gene involved in controlling telomere length in *S. cerevisiae*, is homologous to the human ataxia telangiectasia gene. *Cell* **82**, 823–829 (1995).
- Dahlen, M., Olsson, T., Kanter-Smoler, G., Ramne, A. & Sunnerhagen, P. Regulation of telomere length by checkpoint genes in *Schizosaccharomyces pombe*. *Mol. Biol. Cell.* **9**, 611–621 (1998).
- Ritchie, K. B., Mallory, J. C. & Petes, T. D. Interactions of TLC1 (which encodes the RNA subunit of telomerase), TEL1, and MEC1 in regulating telomere length in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 6065–6075 (1999).
- Anderson, P. Mutagenesis. *Methods Cell Biol.* **48**, 31–58 (1995).
- Hodgkin, J. A., Horvitz, H. R. & Brenner, S. Nondisjunction mutants of the nematode *C. elegans*. *Genetics* **91**, 67–94 (1979).
- Plasterk, R. H. A. & van Leunen, H. G. A. M. *Transposons in C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 97–116 (Cold Spring Harbor Laboratory Press, Plainview, 1997).
- Herman, R. K., Kari, C. K. & Hartman, P. S. Dominant X-chromosome nondisjunction mutants of *Caenorhabditis elegans*. *Genetics* **102**, 379–400 (1982).

25. McClintock, B. The stability of broken ends of chromosomes of *Zea mays*. *Genetics* **26**, 234–282 (1941).
26. Albertson, D. G. & Thomson, J. N. Segregation of holocentric chromosomes at meiosis in the nematode, *Caenorhabditis elegans*. *Chromosome Res.* **1**, 15–26 (1993).
27. Wicky, C. *et al.* Telomeric repeats (TTAGGC)_n are sufficient for chromosome capping function in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **93**, 8983–8988 (1996).
28. van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).
29. d'Adda di Fagagna, F. *et al.* Functions of poly(ADP-ribose) polymerase in controlling telomere length and chromosomal stability. *Nature Genet.* **23**, 76–80 (1999).
30. Lundblad, V. & Blackburn, E. H. An alternative pathway for yeast telomere maintenance rescues *est1*–senescence. *Cell* **73**, 347–360 (1993).
31. Dean, F. B., Lian, L. & O'Donnell, M. cDNA cloning and gene mapping of human homologs for *Schizosaccharomyces pombe rad17*, *rad1*, and *hus1* and cloning of homologs from mouse, *Caenorhabditis elegans*, and *Drosophila melanogaster*. *Genomics* **54**, 424–436 (1998).
32. Bluyssen, H. A. *et al.* A human and mouse homolog of the *Schizosaccharomyces pombe rad1+* cell cycle checkpoint control gene. *Genomics* **54**, 331–337 (1998).
33. al-Khodairy, F. & Carr, A. M. DNA repair mutants defining G2 checkpoint pathways in *Schizosaccharomyces pombe*. *EMBO J.* **11**, 1343–1350 (1992).
34. Marathi, U. K. *et al.* RAD1, a human structural homolog of the *Schizosaccharomyces pombe* RAD1 cell cycle checkpoint gene. *Genomics* **54**, 344–347 (1998).
35. Freire, R. *et al.* Human and mouse homologs of *Schizosaccharomyces pombe rad1(+)* and *Saccharomyces cerevisiae RAD17*: linkage to checkpoint control and mammalian meiosis. *Genes Dev.* **12**, 2560–2573 (1998).
36. Weinert, T. A. & Hartwell, L. H. Cell cycle arrest of cdc mutants and specificity of the RAD9 checkpoint. *Genetics* **134**, 63–80 (1993).
37. Enoch, T., Carr, A. M. & Nurse, P. Fission yeast genes involved in coupling mitosis to completion of DNA replication. *Genes Dev.* **6**, 2035–2046 (1992).
38. Morrow, D. M., Tagle, D. A., Shiloh, Y., Collins, F. S. & Hieter, P. *TEL1*, an *S. cerevisiae* homolog of the human gene mutated in ataxia telangiectasia, is functionally related to the yeast checkpoint gene *MEC1*. *Cell* **82**, 831–840 (1995).
39. Matsuur, A., Naito, T. & Ishikawa, F. Genetic control of telomere integrity in *Schizosaccharomyces pombe*, *rad3(+)* and *tel1(+)* are parts of two regulatory networks independent of the downstream protein kinases *chk1(+)* and *cds1(+)*. *Genetics* **152**, 1501–1512 (1999).
40. Sandell, L. L. & Zakian, V. A. Loss of a yeast telomere: arrest, recovery, and chromosome loss. *Cell* **75**, 729–739 (1993).
41. Henderson, E. in *Telomeres* (eds Blackburn, E. H. & Greider, C. W.) 35–68 (Cold Spring Harbor Laboratory Press, Plainview, 1995).
42. Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
43. Lydall, D. & Weinert, T. Yeast checkpoint genes in DNA damage processing: implications for repair and arrest. *Science* **270**, 1488–1491 (1995).
44. Wellinger, R. J., Wolf, A. J. & Zakian, V. A. *Saccharomyces* telomeres acquire single-strand TG1-3 tails late in S phase. *Cell* **72**, 51–60 (1993).
45. Nugent, C. L., Hughes, T. R., Lue, N. F. & Lundblad, V. Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science* **274**, 249–252 (1996).
46. Mills, K. D., Sinclair, D. A. & Guarente, L. *MEC1*-dependent redistribution of the Sir3 silencing protein from telomeres to DNA double-strand breaks. *Cell* **97**, 609–620 (1999).
47. Martin, S. G., Laroche, T., Suka, N., Grunstein, M. & Gasser, S. M. Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. *Cell* **97**, 621–633 (1999).
48. Fang, G. & Cech, T. R. Telomerase RNA localized in the replication band and spherical subnuclear organelles in hypotrichous ciliates. *J. Cell Biol.* **130**, 243–253 (1995).
49. Chin, L. *et al.* p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. *Cell* **97**, 527–538 (1999).
50. Sulston, J. & Hodgkin, J. in *The Nematode Caenorhabditis elegans* (ed. Wood, W. B.) 587–606 (Cold Spring Harbor Laboratory Press, Plainview, 1988).

Acknowledgements

We thank R. Hill, R. H. A. Plasterk and P. Kuwabara for advice and discussions; K. Van Auken for *nob-1* (*ct230*); F. Müller and A. Coulson for plasmids; and A. Gartner, K. J. Patel, R. Harris, M. O'Reilly, D. Rhodes, V. P. C. C. Yu, A. Woollard, N. Hopper, M. Bickle, S. Sokol, and C. Milne for discussions and comments on this manuscript. Some strains were from the *C. elegans* Genetics Center (St. Paul, Minnesota), which is supported by the National Center for Research Resources. This work was supported by the Medical Research Council UK and the Howard Hughes Medical Institute. S.A. is a recipient of a Burroughs Wellcome Fund Hitchings-Elion Fellowship.

Correspondence and requests for materials should be addressed to S.A. (e-mail: shawn@mrc-lmb.cam.ac.uk). The sequences can be found in the GenBank under accession codes: AF073524 (hRAD1); AF074718 (mRAD1); AF124501 (dmRAD1); AF076843 (MRT-2); P22193 (spRad1p); p48581 (scRad17p); and P14746 (umRec1).

A reduced estimate of the number of kilometre-sized near-Earth asteroids

David Rabinowitz*, Eleanor Helin†, Kenneth Lawrence†
& Steven Pravdo†

* Yale University Physics Department, 266 Whitney Avenue, JWG 552, New Haven, Connecticut 06511, USA

† Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA

Near-Earth asteroids are small (diameters < 10 km), rocky bodies with orbits that approach that of the Earth (they come within 1.3 AU of the Sun). Most have a chance of approximately 0.5% of colliding with the Earth in the next million years. The total number of such bodies with diameters > 1 km has been estimated to be in the range 1,000–2,000, which translates to an approximately 1% chance of a catastrophic collision with the Earth in the next millennium^{1,2}. These numbers are, however, poorly constrained because of the limitations of previous searches using photographic plates. (One kilometre is below the size of a body whose impact on the Earth would produce global effects³.) Here we report an analysis of our survey for near-Earth asteroids that uses improved detection technologies. We find that the total number of asteroids with diameters > 1 km is about half the earlier estimates. At the current rate of discovery of near-Earth asteroids, 90% will probably have been detected within the next 20 years.

Our analysis is based on the results of the Near-Earth Asteroid Tracking (NEAT) programme of the Jet Propulsion Laboratory of the California Institute of Technology⁴. It is an automated search for NEAs that uses a large-format, charge-coupled device (CCD) and a 1.0-m aperture telescope at a US Air Force facility on the summit of Haleakala Crater on the Hawaiian island of Maui. Six years after the first CCD discovery of an NEA by the Spacewatch Telescope in 1989 (ref. 5), NEAT became the second search programme to use computer software to detect NEAs in CCD images. Additional searches employing CCDs are now in operation³. Compared to previous photographic methods, which required a trained observer to identify asteroids by visual inspection of developed films^{6,7}, these automated searches are much more systematic. There is no variation in detection efficiency as a result of the varying abilities of the observers. Furthermore, the automated searches provide a record of every detectable asteroid, including those in the main belt, which appear $\sim 1,000$ times more frequently than NEAs. The detection frequency of these additional objects, which was largely unmeasured in previous photographic searches, provides a good indication of search efficiency. The new searches thus allow a more accurate determination of the size of the NEA population.

In the period March 1996 to August 1998, NEAT searched $\sim 35,000$ square degrees of sky to limiting visual magnitude $V = 19.1 \pm 0.1$ and detected 45 NEAs (see Table 1). Of these, 26 have absolute magnitudes $H < 18$; this value corresponds to the limiting diameter $d > 1$ km, assuming an albedo of 0.10, which is close to the

mean albedo for larger main-belt asteroids. Previous population estimates have assumed the same or nearly the same H limit^{1,2,7,8}. This is a reasonable assumption even if the albedo distribution is bimodal, with comparable numbers of NEAs of both low and high albedo. For the purpose of comparison with previous works, we preserve the assumption.

To estimate the size of the total NEA population from this incomplete sampling, we use a computer program to simulate the detection by NEAT of a large, fictitious population of NEAs with random, but realistically distributed orbits. Given the known position and time of observation for each field surveyed by NEAT, and given the known detection efficiencies of NEAT (ref. 4) as a function of V and angular rate, ω , the program determines the H dependence of our net detection efficiency (that is, the fractional completeness of our survey). Dividing the H distribution of our detections by our net efficiency, we thus obtain the total number of NEAs, $N(H)$, as a function of H .

A similar computer simulation was used previously to determine $N(H)$ based on the observations of the Spacewatch telescope^{9,10}.

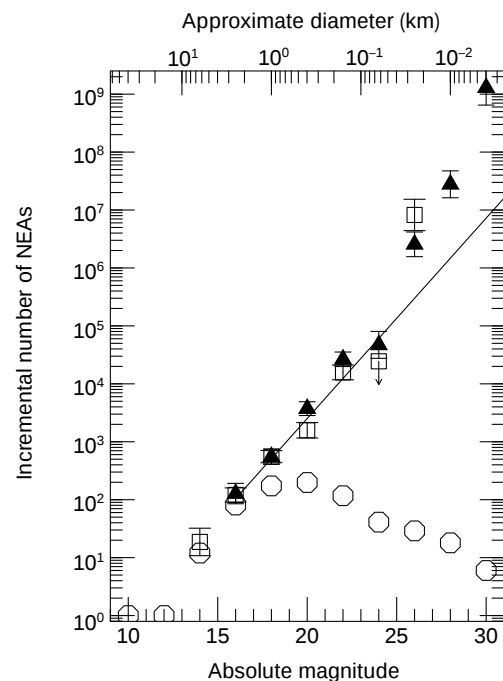


Figure 1 Number of near-Earth asteroids versus absolute magnitude and diameter. Each point shows the number (N) counted within the preceding two-magnitude interval. Squares and triangles show the complete population determined from bias corrections to the NEAT (Near-Earth Asteroid Tracking) and Spacewatch surveys, respectively. Also shown is the total observed population (octagons) as of June 1999. Error bars indicate only the counting uncertainties of the surveys. An upper limit assuming a single detection is shown for absolute magnitude $H = 22$ – 24 where NEAT detected no NEAs. The previously unnormalized Spacewatch population is scaled to best overlap the NEAT result. Additional systematic errors are discussed in the text. The solid line is a linear fit to the NEAT data with a slope of 0.35, assuming $\log [N]$ is proportional to absolute magnitude over the range 16–22. The diameter scale is drawn assuming all NEAs have albedo 0.10. The NEAT distribution was derived assuming a slope parameter $G = 0.15$ to describe the dependence of visual magnitude V on solar phase angle (see definition in ref. 14) and assuming a limiting magnitude for the NEAT survey that is brighter by 0.1 magnitudes than reported in ref. 4. This correction, which accounts for the average effects of variable extinction, seeing, and telescope tracking errors, was determined from the dispersion between measured and predicted magnitudes for catalogue standards appearing in each survey field. Bias corrections to both the NEAT and Spacewatch surveys assume a random distribution of NEA orbits, weighted by their representation in a debiased distribution^{2,10} determined from photographic detections of kilometre-sized NEAs by the Palomar Comet and Asteroid Survey^{6,7}.

Table 1 Number of detected NEAs versus absolute magnitude, H

H range	Number
12–14	2
14–16	10
16–18	14
18–20	8
20–22	9
22–24	0
24–26	2

Only NEAT detections made from March 1996 to August 1998 are shown. Some of these were observations of previously known NEAs, but in all cases the detections were incidental — there being no knowledge that the NEAs would appear in the survey fields.

Owing to uncertainties in the absolute detection efficiencies of the telescope, however, the resulting distribution was not normalized. Only the relative number of NEAs as a function of H was determined. To prevent duplication of any unknown computational errors in the Spacewatch analysis, we do not incorporate any of the software from that analysis into our new program. We also use new algorithms to predict the position versus time of the NEAs and to model the search pattern. In order to obtain comparable results, however, we preserve the method used to assign fictitious NEA orbits.

Figure 1 shows incremental values of $N(H)$ for the total observed NEA population, and for the complete population determined both from our own analysis of the NEAT data and from the earlier analysis of the Spacewatch data. Comparing the NEAT and Spacewatch results, it is apparent that the measurement of $N(H)$ is repeatable. Of the six magnitude ranges from $H = 14$ –26 where both curves extend, the curves deviate significantly only from $H = 18$ –20. A linear fit to the NEAT curve in the range $H = 16$ –22 is consistent with both data sets.

This overall correspondence cannot easily be attributed to selection effects or analysis errors. Differences in the instrumentation, survey pattern, area coverage, and the dependence of search efficiency on V and ω would lead to different selection effects for the two programs. If these were insufficiently accounted for, or if there were computational errors in the analyses, the resulting curves would have been affected differently.

The consistency between the NEAT distribution and the known population for $H < 16$ ($d > \sim 5$ km) provides additional support for our derivation. Because the known population of NEAs with $H < 16$ is nearly complete (10 of the 12 NEAs with $H < 16$ detected by NEAT were previously known, thus indicating a completeness fraction of $\sim 80\%$), any significant discrepancy would have indicated an error in the analysis.

To determine the magnitude of possible systematic errors, we repeated our simulation assuming lower and higher values for the limiting magnitude ($V = 18.9$ and $V = 19.1$) and for the phase parameter ($G = -0.03$ and $G = 0.40$), thus covering the range of uncertainty for these parameters. In all cases, the resulting $N(H)$ curves were similar to our nominal result, with the slope from $H = 14$ –24 changing by less than 5%, and the value at $H = 18$ changing by less than 14%.

To account for possibly uncorrected or improperly corrected observational biases in our assumed orbit distribution, we repeated the simulation once more—this time choosing random orbits from the total observed population of NEAs with $H < 15$ (41 bodies as of July 1999, ref. 11). Since these larger NEAs have been surveyed to near completion, their orbits are unbiased. On the other hand, their representation of all the possible NEA orbits is sparse. Hence, any deviations from our nominal derivation for $N(H)$ represent an upper limit to the possible errors. On the basis of this comparison, we determine a systematic error of less than 3% in the average slope of $N(H)$, and less than 16% in the value at $H = 18$.

Taking into account both random and systematic errors, we thus conclude that there are 700 ± 230 NEAs with $H < 18$. This number is about half the value of previous estimates; however, given both the old and new uncertainties, we are marginally consistent with previous lower limits. This decrease of a factor of two does not substantially alter the significance of the NEA hazard. We have shown, however, that continuing, automated searches provide a reliable means to characterize the size of the NEA population. We can now confidently predict the level of effort required to survey completely those NEAs capable of global devastation.

Previous survey simulations show that the number of NEAs remaining undiscovered will diminish nearly exponentially with time, after an initial fast decay, if the detection rate of both known and unknown NEAs is held constant^{8,12}. For NEAs with high and low albedos similar to the S- and C-type asteroids that dominate the

main belt, limit $d > 1$ km corresponds to limits $H < 17.5$ and $H < 19.0$, respectively. Given the current yearly detection rates of 50 and 110 at these limits (as of the 12-month period ending June 1999, see ref. 11), and given our own estimates for the respective completeness levels of the known population ($51 \pm 17\%$ and $26 \pm 8\%$), it will take 7 ± 4 and 22 ± 14 years to reach 90% completion for these two types. Since the NEAs have a broad range of spectral types for which the associated albedos are largely spanned by C and S albedos¹³, the time to 90% completion for all NEAs larger than 1 km is probably close to the average for these two types, or 15 ± 10 years. Doubling the current world-wide detection rate would therefore lead to near completion in the next decade. □

Received 13 August; accepted 17 November 1999.

1. Shoemaker, E. M., Wolfe, R. F. & Shoemaker, C. S. in *Global Catastrophes in Earth History* (eds Sharpton, V. L. & Ward, P. D.) 155–170 (Geological Society of America, Boulder, Colorado, 1990).
2. Rabinowitz, D. L., Bowell, E., Shoemaker, E. M. & Muinonen, K. in *Hazards due to Comets and Asteroids* (ed. Gehrels, T.) 285–312 (Univ. Arizona Press, Tucson, 1994).
3. The Threat and the Opportunity of Asteroid and other Near-Earth Objects: Hearing before the Subcommittee on Space and Aeronautics of the Committee on Science (US House of Representatives) (US Government Printing Office, Washington DC, 1998).
4. Pravdo, S. H. *et al.* The near-Earth asteroid tracking (NEAT) program: an automated system for telescope control, wide-field imaging, and object detection. *Astron. J.* **117**, 1616–1633 (1999).
5. Gehrels, T. Scanning with charge-coupled devices. *Space Sci. Rev.* **58**, 347–375 (1991).
6. Helin, E. F. & Dunbar, R. S. Search techniques for near-Earth asteroids. *Vistas Astron.* **33**, 21–37 (1990).
7. Helin, E. F. & Shoemaker, E. M. The Palomar planet-crossing asteroid survey, 1973–1978. *Icarus* **40**, 321–328 (1979).
8. Harris, A. W. in *Collisional Processes in the Solar System* (eds Rickman, H. & Marov, M.) (Kluwer, Dordrecht, in the press).
9. Rabinowitz, D. L. The size distribution of the Earth-approaching asteroids. *Astrophys. J.* **407**, 412–427 (1993).
10. Rabinowitz, D. L. The size and shape of the near-Earth asteroid belt. *Icarus* **111**, 364–377 (1994).
11. Marsden, B. G. Electronic Listing at the Minor Planet Center of the IAU (cited July '99) (<http://cfa-www.harvard.edu/iau/NEO/TheNEOPage.html>).
12. Bowell, E. & Muinonen, K. in *Hazards due to Comets and Asteroids* (ed. Gehrels, T.) 149–197 (Univ. Arizona Press, Tucson, 1994).
13. McFadden, L.-A., Tholen, D. J. & Veeder, G. J. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 442–467 (Univ. Arizona Press, Tucson, 1989).
14. Bowell, E. *et al.* in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 524–556 (Univ. Arizona Press, Tucson, 1989).

Acknowledgements

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Correspondence and requests for materials should be addressed to D. R. (e-mail: david.rabinowitz@yale.edu).

Measurement of the spatial coherence of a trapped Bose gas at the phase transition

I. Bloch, T. W. Hänsch & T. Esslinger

Sektion Physik, Ludwig-Maximilians-Universität, Schellingstrasse 4/III, D-80799 Munich, Germany, & The Max-Planck-Institut für Quantenoptik, D-85748 Garching, Germany

The experimental realization of Bose–Einstein condensates of dilute gases^{1–3} has allowed investigations of fundamental concepts in quantum mechanics at ultra-low temperatures, such as wave-like behaviour and interference phenomena. The formation of an interference pattern depends fundamentally on the phase coherence of a system; the latter may be quantified by the spatial correlation function. Phase coherence over a long range^{4–7} is the essential factor underlying Bose–Einstein condensation and

related macroscopic quantum phenomena, such as superconductivity and superfluidity. Here we report a direct measurement of the phase coherence properties of a weakly interacting Bose gas of rubidium atoms. Effectively, we create a double slit for magnetically trapped atoms using a radio wave field with two frequency components. The correlation function of the system is determined by evaluating the interference pattern of two matter waves originating from the spatially separated 'slit' regions of the trapped gas. Above the critical temperature for Bose–Einstein condensation, the correlation function shows a rapid gaussian decay, as expected for a thermal gas. Below the critical temperature, the correlation function has a different shape: a slow decay towards a plateau is observed, indicating the long-range phase coherence of the condensate fraction.

So far, experiments with condensed Bose gases have focused on the coherence properties of almost pure condensates, prepared at temperatures well below the critical temperature. First-order coherence of a Bose condensed gas was qualitatively shown by observing an interference pattern when two independent condensates were brought to overlap⁸. Using a spectroscopic method it has also been possible to show that the coherence length of the condensate is equal to its size^{9,10}. Aspects of second-order¹¹ and third-order coherence¹² of trapped condensates and the evolution of the relative phase between two condensates in different spin states were also explored¹³.

Here we probe the relative phase of the trapped gas between two spatially separated 'slit' regions. Each virtual slit region is selected by the frequency of a weak radio wave field that induces transitions between magnetically trapped and untrapped atomic states. The resulting output from the trap consists of two matter-wave beams with different energies. As a result of gravity the beams are collimated and propagate downwards. If the trapped atoms have a coherence length that spans both slit regions, a matter-wave interference pattern is observed in a snapshot of the spatial distribution of the emitted output beam, as shown in Fig. 1. The degree of spatial coherence of the trapped gas is deduced from the fringe contrast of the interference pattern, which is measured both as a function of temperature and slit separation.

Let us first consider the output wave emitted from a magnetically trapped Bose–Einstein condensate when a radio wave with a single frequency ω is applied¹⁴. The magnetic field $B(\mathbf{r})$ of our trap¹⁵ gives

rise to a harmonic potential which confines the condensate into the shape of a cigar, with its long axis oriented perpendicular to the gravitational force. Because of gravity, the minimum of the trapping potential is displaced relative to the minimum of the magnetic field ($13\ \mu\text{m}$ for our trapping parameters with a pure condensate having a radial diameter of $\sim 10\ \mu\text{m}$). Atoms are therefore predominantly transferred into the untrapped state at the intersection of the condensate with the shell that is determined by the electron spin resonance condition $1/2\mu_B|B(\mathbf{r})| = \hbar\omega$, with μ_B being the Bohr magneton¹⁴. Over the size of the trapped atomic cloud the height at which the resonance condition is fulfilled varies only by a small amount. The experimental situation is therefore well described by a one-dimensional model: in the region of the condensate the output waves can be approximated by Airy functions $\text{Ai}(\zeta)$, which are the eigenfunctions of massive particles in the gravitational potential¹⁶. These functions are characterized by an energy $E = mgz_0$, where z_0 is

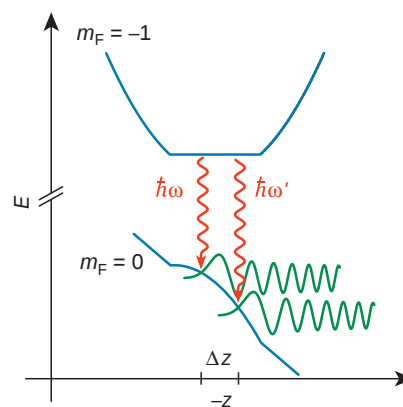


Figure 2 Measurement of the spatial coherence of a trapped Bose–Einstein condensate. A radio-wave field with two frequency components (red lines) transfers magnetically trapped atoms ($m_F = -1$) into untrapped states ($m_F = 0$). Two matter wave beams (green lines) are emitted from spatially separated locations in the trap, thereby effectively forming a double slit. The beams are accelerated by gravity and overlap outside the trap so that an interference pattern can be observed. A high contrast is detected if the trapped gas is spatially coherent over both slit regions. The harmonic potential for the trapped state and the linear potential for the untrapped state are modified by the mean field of the condensate (blue lines).

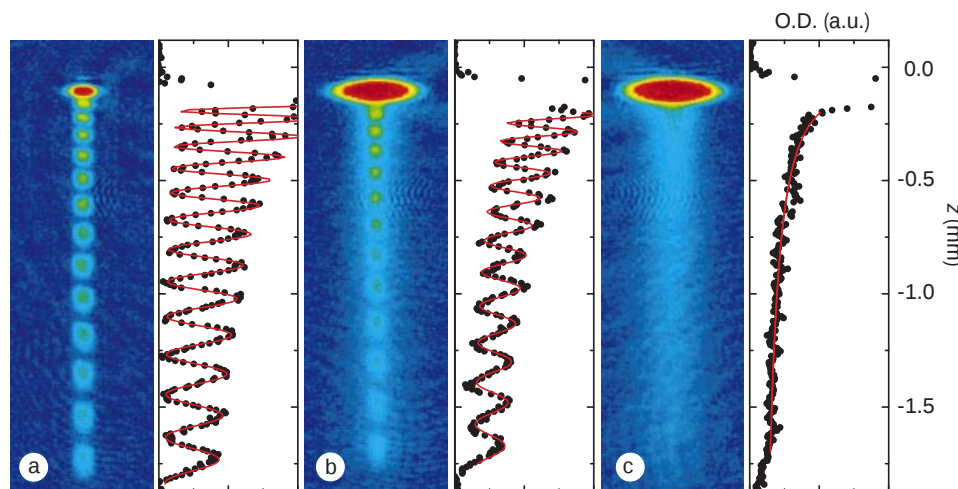


Figure 1 Interference pattern of matter-wave beams emitted from two spatially separated regions of a trapped Bose gas. **a–c**, The absorption images show the spatial distribution of the output beams for three different temperatures in false colours. For a temperature well below the critical temperature T_c the spatially uniform phase of the condensate results in a high-contrast interference pattern (**a**). When the temperature is increased to just below T_c the contrast of the interference pattern is reduced (**b**), and it vanishes

completely for temperatures above T_c (**c**). Each of the images has a size of $0.6\ \text{mm} \times 2\ \text{mm}$. The frequency difference between the two components of the radio-wave field was $1,000\ \text{Hz}$, which corresponds to a slit separation of $465\ \text{nm}$. The plots next to the images show $21\text{-}\mu\text{m}$ -wide vertical cuts through the centres of the absorption images. The horizontal axes of the plots show the optical densities. We obtain the visibilities V by using equation (1) to fit the data.

the apex of the corresponding classical trajectory, m the atomic mass and g the gravitational acceleration. The scaled parameter ζ is given by $\zeta = (z - E/mg)/l$, with a natural length scale $l = (\hbar^2/2m^2g)^{1/3}$. Outside the condensate region the asymptotic behaviour of the outgoing matter wave ϕ_{out} is given by¹⁶: $\phi_{\text{out}}(\zeta) \propto |\zeta|^{-1/4} \exp(i(2/3)|\zeta|^{3/2})$. An output wave of sharply defined energy is produced if the coupling into the untrapped state is weak and sustained for a long enough period¹⁷. Such output coupling can be described by a rate proportional to the square of the overlap integral between the output state and the trapped-state wavefunction, where the maximum contribution to the integral is localized in the region determined by the resonance condition. For our trapping parameters the mean-field interaction energy of the condensate causes only a slight distortion in the gravitational potential so that Airy functions are a good approximation for the output waves.

The radio-wave field applied in the experiment is composed of two frequencies ω and ω' , coupling the trapped condensate to two outgoing waves which have a difference in energy of $\Delta E = \hbar(\omega - \omega')$. This results in a spatial separation $\Delta z = \Delta E/mg$ between the classical turning points of the corresponding Airy functions, as illustrated in Fig. 2. Using the asymptotic approximation¹⁶, the atomic density distribution n_{out} of the output waves is given by:

$$n_{\text{out}}(z) = |\phi_{\text{out}}(\zeta) + \phi_{\text{out}}(\zeta')|^2 \propto \frac{1}{\sqrt{|z|}} \left\{ 2 + 2V \cos(q\sqrt{|z|} + (\omega - \omega')t) \right\} \quad (1)$$

Here the variable q is given by $q = m\Delta z\sqrt{2g}/\hbar$. The same interference term is obtained when two point sources for atomic de Broglie waves are considered that are positioned in the gravitational potential with a distance Δz between them¹⁸. We have additionally introduced the visibility V , which is equal to one for the case of perfect coherence and equal intensities of the two waves.

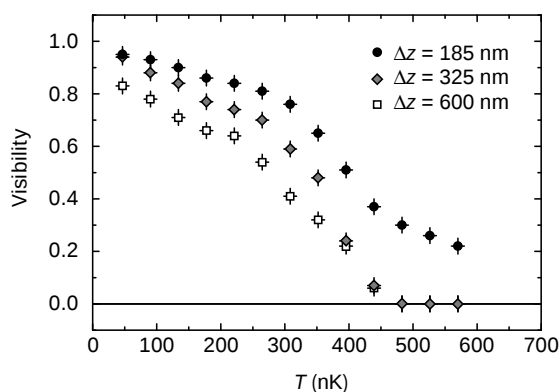


Figure 3 Spatial coherence of the Bose gas above and below the transition temperature T_c . The temperature dependency of the measured visibility is plotted for three different slit separations. Above T_c (≈ 430 nK) the visibility is zero for slit separations of 600 nm (white squares) and 325 nm (grey diamonds). A sudden increase in the visibility is detected when the temperature drops below T_c . Close to $T = 0$ the visibility reaches a maximum value of 0.95. The visibility has a qualitatively different behaviour for a slit separation of 185 nm (black circles). The visibility does not vanish at temperatures above T_c and the increase in visibility below the transition temperature is less pronounced. This is because the coherence length of the trapped gas at the transition temperature is larger than the slit separation. The slit separations of 185 nm, 325 nm and 600 nm correspond to differences of 400 Hz, 700 Hz and 1,300 Hz between the two frequency components of the radio-wave field. With increasing slit separation the fringe spacing in the interference pattern becomes smaller (see equation (1)) and the visibility is reduced by the resolution limit of our imaging system, which is $8 \mu\text{m}$. The data points displayed in Fig. 3 do not take this into account. Therefore the visibility appears to be lower for large-slit separations even for almost pure condensates at $T = 50$ nK.

The visibility becomes zero if the contrast of the observed interference pattern vanishes.

We produce samples of cold Bose gases using the same experimental set-up and a similar procedure as described in our previous work^{14,15}. In brief, rubidium-87 atoms are trapped and cooled in a magneto-optical trap. Then the atoms are optically pumped into the hyperfine ground state with the total angular momentum $F = 1$ and the magnetic quantum number $m_F = -1$. In this state the atoms are magnetically trapped and further cooled to the desired final temperature by radio-frequency-induced evaporation. A very compact magnetic trap is employed which combines the quadrupole with the Ioffe geometry¹⁵. It is placed inside a magnetic shield enclosure so that field fluctuations due to the environment are reduced to a level below 0.1 mG. The radial and axial trapping frequencies are $\omega_{\perp} = 2\pi \times 140$ Hz and $\omega_{\text{ax}} = 2\pi \times 13$ Hz, respectively. The trapped clouds we produce typically contain four million atoms at the transition temperature. To determine the temperature of the gas we apply the usual time-of-flight technique¹⁹. We estimate the absolute error of our temperature measurements to be 20 nK. The shot-to-shot fluctuations in the temperature are below this value.

The matter-wave beams are extracted from the condensate over a period of 13 ms, using a radio-wave field with two frequency components ω and ω' of the same amplitude ($0.4 \text{ mG} \pm 0.1 \text{ mG}$). After a time delay of 2 ms the magnetic trapping field is switched off and the absorption image is taken another 3 ms later. We obtain the fringe visibility V from a fit of the absorption image using equation (1), as illustrated in Fig. 1. In the experiments described below we adjusted the radio frequencies ω and ω' in such a way that the slit regions are centred around the trap minimum. Within a range of 4 kHz a simultaneous shift of both radio frequencies by the same amount results in a change in the measured visibility by less than 5%.

The fringe visibility as a function of temperature is shown in Fig. 3. Above the critical temperature T_c the gas has a coherence

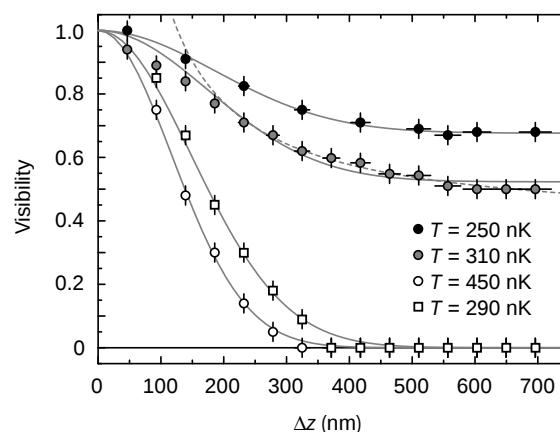


Figure 4 Spatial correlation function of a trapped Bose gas. The fringe visibility is measured as a function of slit separation for temperatures above and below the critical temperature T_c . The white circles and white squares show the measurements for thermal gases at temperatures of 450 nK and 290 nK, respectively. (To prepare a thermal gas at a temperature of 290 nK we reduced the number of atoms in the trap to 10^6). The grey data points (310 nK) and the black data points (250 nK) are the results obtained for temperatures below T_c , where the visibility decays to a nonzero value due to the long-range phase coherence of the condensate fraction. The decay of the visibility shows a longer tail than would be expected for the corresponding de Broglie wavelength of the thermal fraction. Both data sets are fitted with a Gaussian function and an offset. For the temperature of 310 nK the last 11 data points are also fitted with a function proportional to $1/\Delta z$ (dashed line) and an offset. The distortion of the gravitational potential due to the mean field of the condensate leads to an uncertainty in the slit separation between 3 nm and 50 nm (see error bars). The data points displayed in the figure are corrected for the reduction in visibility which is due to the limited resolution of our imaging system.

length that is shorter than the slit separation, which was 325 nm in this case. A sudden increase in the visibility is observed when the gas is cooled below the critical temperature. For very low temperatures the visibility reaches a value of almost one. The same characteristics are observed for a slit separation of 600 nm. Qualitatively different behaviour can be seen when the slit separation is reduced to 185 nm. The slit separation is then smaller than the coherence length of the thermal gas and the interference pattern is still visible above T_c .

A more detailed physical picture is obtained from the close relation between the measured visibility V and the first-order correlation function, which is expressed by the density matrix $\rho(\mathbf{r}, \mathbf{r}')$ of the system and characterises correlations between the matter-wave field at points $\mathbf{r}$ and $\mathbf{r}'$:

$$\rho(\mathbf{r}, \mathbf{r}') = \langle \hat{\Psi}^\dagger(\mathbf{r}) \hat{\Psi}(\mathbf{r}') \rangle = \psi^*(\mathbf{r}) \psi(\mathbf{r}') + \langle \delta \hat{\Psi}^\dagger(\mathbf{r}) \delta \hat{\Psi}(\mathbf{r}') \rangle \quad (2)$$

On the right hand side of the equation the quantized boson field operator $\hat{\Psi}(\mathbf{r})$ has been decomposed into the condensate wavefunction $\psi(\mathbf{r}) = \langle \hat{\Psi}(\mathbf{r}) \rangle$ and fluctuating part $\delta \hat{\Psi}(\mathbf{r})$. The condensate wavefunction gives rise to a long-range phase coherence and the fluctuating part describes the correlations in the excitation spectrum. For a trapped ideal gas at finite temperature the correlation function takes the form:

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_k N_k u_k^*(\mathbf{r}) u_k(\mathbf{r}') \quad (3)$$

Here N_k denotes the population of the energy eigenstates u_k with energy E_k (ref. 20). The ground state describing the condensate is included in the sum.

The density distribution n_{out} of the output matter waves can be expressed in terms of the population of the trap eigenstates:

$$n_{\text{out}}(z) \propto \frac{1}{\sqrt{|z|}} \times \left\{ \sum_k N_k (|\tilde{u}_k|^2 + |\tilde{u}'_k|^2) + 2 \cos(q\sqrt{|z|} + (\omega - \omega')t) \sum_k N_k \tilde{u}_k \tilde{u}'_k \right\} \quad (4)$$

The overlap integrals between the trapped and output states, which describe the coupling induced by the two frequencies ω and ω' , are given by $\tilde{u}_k = \int \text{Ai}(\zeta_k) u_k(z) dz$ and $\tilde{u}'_k = \int \text{Ai}(\zeta'_k) u_k(z) dz$. The index k of the scaled parameters ζ_k and ζ'_k indicates that the energies of the Airy functions are given by $E_k - \hbar\omega$ and $E_k - \hbar\omega'$, respectively. The main contributions to the overlap integrals are localized at the slit regions, where the trapped atoms are resonantly transferred into untrapped states. The observed visibility V (see equation (1)) can therefore be related to the density matrix and the correlation function by

$$V = \frac{2 \sum_k N_k \tilde{u}_k \tilde{u}'_k}{\sum_k N_k [|\tilde{u}_k|^2 + |\tilde{u}'_k|^2]} \approx \frac{\rho(z, z')}{\rho(z, z)} \quad (5)$$

where it is assumed that the density of the trapped gas varies negligibly between the slit regions z and z' . In a numerical calculation we have compared the correlation function of a thermally populated one-dimensional harmonic oscillator with the expression for the visibility given in equation (5). We found excellent agreement for the experimental parameters, as the length scale $\sqrt{\hbar/m\omega_\perp}$ of the radial harmonic oscillator ground state is three times larger than the length scale l of the Airy function.

The fluctuating term in the correlation function of an interacting Bose gas at finite temperatures can be expressed in terms of the quasi-particle excitations of the condensed gas²¹. In this case the energy spectrum of the harmonic oscillator in equation (4) is replaced by the quasi-particle excitation spectrum and the overlap integrals are calculated using the mode functions of the quasi-particle states.

In a second series of experiments we have measured the fringe visibility as a function of slit separation in order to determine the correlation function of the trapped Bose gas. The data sets are displayed in Fig. 4 and show a clear distinction between condensed and non-condensed gases. For temperatures above T_c the visibility decays to zero as the slit separation increases, starting from an initial value which is close to one. We have fitted each set of data points using a gaussian curve $V = \exp(-\pi \Delta z^2/w^2)$ and obtained widths w of 294 ± 6 nm and 372 ± 8 nm. Both values are $\sim 5\%$ larger than the corresponding thermal de Broglie wavelengths λ_T of the gas, which are 278 ± 6 nm and 346 ± 12 nm. A classical ideal gas should have a correlation function of gaussian shape, with a width w given by the thermal de Broglie wavelength $\lambda_T = \sqrt{2\pi\hbar^2/(mk_B T)}$, where k_B is Boltzmann's constant. The properties of the Bose–Einstein statistics enhance the occupation of the low-energy states in a non-condensed gas resulting in a slower decay of the correlation function²⁰, as measured in our experiments.

Below the transition temperature T_c , the correlation function is modified in two significant ways. First, the correlation function decays from an initial value near unity to a plateau. The decay characterises the contribution of the non-condensed fraction to the correlation function and the visibility at the plateau is a direct measure of the condensate fraction in the slit region. At the plateau we have measured a visibility of 0.52 ± 0.03 for a temperature of 310 nK and 0.68 ± 0.03 for a temperature of 250 nK. Second, the coherence length of the non-condensed fraction is found to be substantially larger than the thermal de Broglie wavelength. We fitted both data sets with gaussian functions plus constant offsets and found widths of 428 ± 26 nm and 463 ± 16 nm for temperatures of 310 nK and 250 nK, respectively. These widths are 28% and 24% larger than the corresponding thermal de Broglie wavelengths, which are 335 ± 11 nm and 373 ± 15 nm. The slower decay of the visibility can be attributed to the long wavelength of the low-lying, phonon-type excitations of the condensed Bose gas. When the slit separation approaches the radial size of the condensate the correlation function should decay to zero, due to the finite coherence length of the condensate^{9,10}.

For a homogenous Bose gas at temperatures below T_c , the correlation function is predicted to decay towards an offset as $1/|\mathbf{r} - \mathbf{r}'|$ (ref. 22). The slit separations chosen for our measurements are small compared to the size of the condensate ground state, so that the observed behaviour should be similar to that of a homogenous system. We fitted the tail of our visibility data with a function proportional to $1/\Delta z$ and found good agreement. Furthermore, our measurements are in good qualitative agreement with numerical calculations presented in ref. 20, where the correlation function for a condensed Bose gas in a trap was calculated.

In a further experiment we have analysed the reproducibility of the interference pattern for a slit separation of 465 nm. For temperatures between 50 nK and the critical temperature we found a reproducible interference pattern, demonstrating that the condensed fraction is not fragmented into independent condensates.

Received 7 September; accepted 15 November 1999.

- Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose–Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
- Davis, K. B. *et al.* Bose–Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
- Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose–Einstein condensation of lithium: Observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997).
- Penrose, O. On the quantum mechanics of helium II. *Phil. Mag.* **42**, 1373–1377 (1951).
- Penrose, O. & Onsager, L. Bose–Einstein condensation and liquid helium. *Phys. Rev.* **104**, 576–584 (1956).
- Beliaev, S. T. Application of the method of quantum field theory to a system of bosons. *J. Exp. Theor. Phys. (USSR)* **34**, 417–432 (1958).
- Yang, C. N. Concept of off-diagonal long-range order and the quantum phases of liquid He and of superconductors. *Rev. Mod. Phys.* **34**, 694–704 (1962).
- Andrews, M. R. *et al.* Observation of interference between two Bose condensates. *Science* **275**, 637–641 (1997).

9. Stenger, J. *et al.* Bragg spectroscopy of a Bose–Einstein condensate. *Phys. Rev. Lett.* **82**, 4569–4573 (1998).
10. Hagley, E. W. *et al.* Measurement of the coherence of a Bose–Einstein condensate. *Phys. Rev. Lett.* **83**, 312–315 (1999).
11. Ketterle, W. & Miesner, H. -J. Coherence properties of Bose–Einstein condensates and atom lasers. *Phys. Rev. A* **56**, 3291–3293 (1997).
12. Burt, E. A. *et al.* Coherence, correlations, and collisions: what one learns about Bose–Einstein condensates from their decay. *Phys. Rev. Lett.* **79**, 337–340 (1997).
13. Hall, D. S., Matthews, M. R., Wieman, C. E. & E. A. Cornell. Measurements of relative phase in two-component Bose–Einstein condensates. *Phys. Rev. Lett.* **81**, 1543–1546 (1998).
14. Bloch, I., Hänsch, T. W. & Esslinger, T. Atom laser with a cw output coupler. *Phys. Rev. Lett.* **82**, 3008–3011 (1999).
15. Esslinger, T., Bloch, I. & Hänsch, T. W. Bose–Einstein condensation in a quadrupole-Ioffe-configuration trap. *Phys. Rev. A* **58**, R2664–R2667 (1998).
16. Flügge, S. *Practical quantum mechanics*. 101–107 (Springer, Heidelberg, 1974).
17. Band, Y. B., Julienne P. S. & Trippenbach, M. Radio-frequency output coupling of the Bose–Einstein condensate for atom lasers. *Phys. Rev. A* **59**, 3823–3831 (1999).
18. Anderson, B. & Kasevich, M. Macroscopic quantum interference from atomic tunnel arrays. *Science* **283**, 1686–1689 (1998).
19. Ensher, J. R., Jin, D. S., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Bose–Einstein condensation in a dilute gas: Measurement of energy and ground-state occupation. *Phys. Rev. Lett.* **77**, 4984–4987 (1996).
20. Naraschewski, M. & R. Glauber. Spatial coherence and density correlations of trapped Bose gases. *Phys. Rev. A* **59**, 4595–4607 (1999).
21. Fetter, A. L. Nonuniform states of an imperfect Bose gas. *Ann. Phys.* **70**, 67–101 (1972).
22. Huang, K. *Statistical Mechanics*. 2nd edn., 304 (Wiley, New York, 1987).

Acknowledgements

We thank J. Schneider, W. Zwerger and D. M. Stamper-Kurn for discussions.

Correspondence and requests for materials should be addressed to T.E. (e-mail: tie@mpq.mpg.de).

A first-order liquid–liquid phase transition in phosphorus

Yoshinori Katayama*, Takeshi Mizutani*, Wataru Utsumi*,
Osamu Shimomura*, Masaaki Yamakata† & Ken-ichi Funakoshi†

* Department of Synchrotron Radiation Research, Japan Atomic Energy Research Institute, 1-1-1 Kouto, Mikazuki, Sayo, Hyogo 679-5148, Japan

† Japan Synchrotron Radiation Research Institute, 1-1-1 Kouto, Mikazuki, Sayo, Hyogo 679-5198, Japan

First-order structural phase transitions are common in crystalline solids, whereas first-order liquid–liquid phase transitions (that is, transitions between two distinct liquid forms with different density and entropy) are exceedingly rare in pure substances^{1–4}. But recent theoretical and experimental studies have shown evidence for such a transition in several materials, including supercooled water^{5–8} and liquid carbon^{9,10}. Here we report an *in situ* X-ray diffraction observation of a liquid–liquid transition in phosphorus, involving an abrupt, pressure-induced structural change between two distinct liquid forms. In addition to a known form of liquid phosphorus—a molecular liquid comprising tetrahedral P₄ molecules—we have found a polymeric form at pressures above 1 GPa. Changing the pressure results in a reversible transformation from the low-pressure molecular form into the high-pressure polymeric form. The transformation is sharp and rapid, occurring within a few minutes over a pressure range of less than 0.02 GPa. During the transformation, the two forms of liquid coexist. These features are strongly suggestive of a first-order liquid–liquid phase transition.

Phosphorus has a number of allotropes in the solid state^{11,12}. White P consists of tetrahedral P₄ molecules, whereas black P is a layered structure; red P is usually amorphous. In spite of the wide

variety of structures, the P atom is invariably threefold coordinated. Each allotrope has a different melting temperature: white P melts at 44 °C while black P and red P melt at around 600 °C (ref. 11). It is known that molten white P at 50 °C consists of P₄ molecules¹³; however, no structural study on liquid P at high temperatures has been performed except for an *ab initio* simulation study, which predicts a polymeric form at high temperatures¹⁴. Recent developments in synchrotron radiation technology and large-volume presses enable us to investigate the structures of liquids under high-pressure and high-temperature conditions^{15–18}.

We have studied the structure of the melt of black P by an *in situ* X-ray diffraction method at SPring-8, a new, third-generation synchrotron facility in Japan. Figure 1 shows the experimental paths in the pressure–temperature phase diagram of black P (refs 19, 20). The melting curve of black P has a maximum around 1 GPa, though melting temperatures of most elements increase monotonically as the pressure increases. In the region where the melting curve has a negative slope, the volume of the liquid is less than that of the solid. Hence, the occurrence of a maximum indicates an anomalous decrease in the volume of liquid with an increase in pressure: it has been related to the presence of different structural species in the liquid²¹. In the present study, X-ray diffraction data on the melt of black P were collected at several pressures below and above the melting-curve maximum.

Figure 2 shows the structure factor, *S*(*Q*), at several pressures, as obtained from X-ray diffraction data (here *Q* is the wavenumber). There is a large difference between *S*(*Q*) below and above 1 GPa. A prominent first peak around 1.4 Å^{−1}, which is characteristic of *S*(*Q*) at low pressures, disappears at high pressures. A new maximum around 2.45 Å^{−1} appears above 1 GPa. On the other hand, *S*(*Q*) at 0.77 GPa is almost identical to that at 0.96 GPa, and *S*(*Q*) at

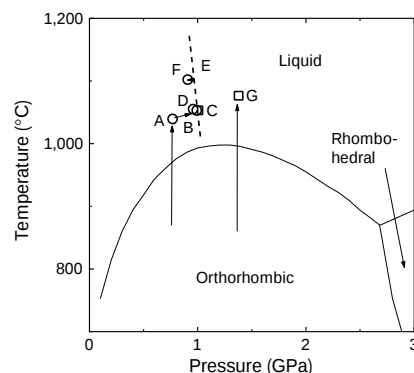


Figure 1 Experimental paths in a previously reported^{19,20} phase diagram of black P. High-pressure and high-temperature conditions were generated using a cubic-type multi-anvil press (SMAP180) installed on BL14B1 at the SPring-8 synchrotron facility. The sample assembly was similar to that used to determine the melting curve of black P (ref. 19). The sample capsule was made of BN, and the temperature was monitored by a thermocouple. When thermocouples broke down under high temperature, temperature was determined by the relation between the electric power of the heater and temperature obtained from a number of experiments under the same conditions. Pressure was determined from the lattice constant of BN (ref. 28). The absolute error in the temperature determination is about ± 50 °C and that of the pressure is about ± 0.1 GPa. Starting materials were either red P or black P. Red P transforms to black P under high pressure and high temperature before it melts. The samples were supplied by Akahama²⁹. The circles (squares) indicate data points where the low-pressure form (high-pressure form) was observed. The low-pressure form was obtained by melting black P at 0.77 GPa (point A). The pressure was then increased and an abrupt change in the diffraction pattern was observed between data points B and C. A reverse change was observed between data points C and D when the pressure was decreased. The high-pressure form was obtained by melting black P above 1 GPa (points E and G). The abrupt change in the diffraction pattern was observed between data points E and F. The dashed line indicates the boundary.

1.01 GPa is similar to that at 1.38 GPa. These results indicate that there are two distinct forms of liquid P. Here, we call them the low-pressure form and the high-pressure form.

Previous studies suggested that the melts of white P, red P and black P are identical, on the basis of their solidification, vapour pressure and density¹¹. In fact, $S(Q)$ for the low-pressure form resembles that for the melt of white P (ref. 13). In $S(Q)$ of the melt of white P at 50 °C, a prominent first peak exists around $1.5 \text{ \AA}^{-1}$ which is attributed to correlation between the P_4 molecules^{13,22}. Thus, it is highly probable that the low-pressure form in our experiments is a molecular liquid. The position of a small first peak in the high-pressure form ($1.25 \text{ \AA}^{-1}$) is different from that of the first peak in the low-pressure form, so that it does not necessarily mean that the low-pressure form remains.

Radial distribution functions, RDFs, derived from $S(Q)$, give further information regarding the local atomic structure (Fig. 3). As there are no density data for liquid P under high pressure, we tentatively assume a density of 2.0 g cm^{-3} for the low-pressure form and 2.8 g cm^{-3} for the high-pressure form in order to calculate the RDF. Owing to the uncertainty in the density, we restrict our discussion to the peak positions and relative peak intensities.

In the RDF for the low-pressure form (upper two curves in Fig. 3), there is a sharp peak with a maximum at $2.22 \text{ \AA}$. The position is in good agreement with the P–P distance in P_4 molecules in the melt of white P, $2.210 \text{ \AA}$ (ref. 13). The intensity in the second-neighbour region (from 2.5 to $4.5 \text{ \AA}$) is relatively low. The high-frequency oscillations are probably spurious due to the narrow Q -range of $S(Q)$. There is a broad third peak around $5.5 \text{ \AA}$. The overall shape of the RDF—with a sharp first peak and a broader, low-intensity second-neighbour region—supports a picture in which the low-pressure form comprises P_4 molecules without strong bonds between them.

The RDF for the high-pressure form has different features (lower two curves in Fig. 3). The most characteristic difference is the appearance of a distinct second peak around $3.5 \text{ \AA}$. The maximum of the first peak shifts to $2.28 \text{ \AA}$. The strong correlation between the second-nearest neighbours is inconsistent with the nature of a molecular liquid, in which molecule–molecule interactions are weak. A possible candidate for the high-pressure form is a polymeric liquid in which P atoms form network structures. In fact, the RDF of the high-pressure form resembles that of red amorphous P, which

has a polymeric nature^{23,24}. In addition, a substantial increase in the intensity of the second peak relative to the first peak in the RDF is also observed in the *ab initio* simulation of the polymerization process in liquid P (ref. 14). Thus, it is very likely that the high-pressure form in our experiments is a polymeric liquid.

The low- and high-pressure forms are obtained by melting black P below and above 1.0 GPa , respectively. In addition, the high-pressure form can be directly obtained from the low-pressure form, and *vice versa*, by a change in pressure. Figure 4 shows the diffraction patterns across the transformation, taken at a diffraction angle of 6.0° . A broad peak around a photon energy of 25 keV in pattern a corresponds to the first maximum of $S(Q)$ for the low-pressure form. The pressure was then increased at a rate of about $0.01 \text{ GPa min}^{-1}$ along path B–C in Fig. 1. During compression, the peak around 25 keV abruptly decreased. When we noticed the change, we tried to keep the pressure constant but the transformation proceeded. Pattern b in Fig. 4 was taken during the transformation. The transformation proceeded during the data acquisition; the relative intensity of the first peak to the second peak continuously changed. After 15 min, the diffraction pattern did not show any further change (pattern c). An increase in pressure of 0.02 GPa results in a completely different diffraction pattern. The reverse change was observed with decreasing pressure along the path C–D. It was hard to see whether or not there was hysteresis in the transition pressure. Patterns d to f were taken with decreasing pressure along the path E–F. The transformation was completed within a pressure range of less than 0.06 GPa and a time interval of 8 min. In subsequent experiments, we noticed that we could change the speed of transformation by a careful control of pressure.

The diffraction pattern taken during the transformation may be perfectly reproduced from the diffraction pattern taken before the transformation and the pattern taken afterwards. For example, adding patterns a and c (in Fig. 4) in the ratio of 38:62 reproduces pattern b (grey line). Pattern e can also be reproduced from patterns d and f. This fact indicates that there is no long-lived intermediate state that contributes to X-ray diffraction; the low-pressure form transforms directly to the high-pressure form. In other words, the low-pressure form and the high-pressure form coexist during the transformation.

The existence of molecular liquid at high temperature is surprising, because the melt of white P rapidly polymerizes and turns into

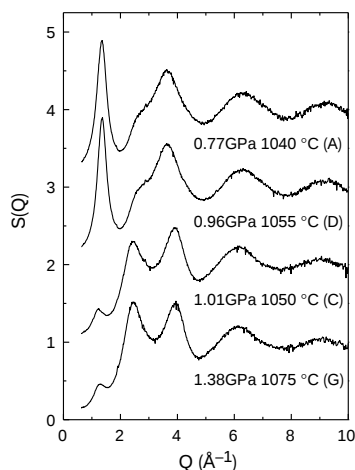


Figure 2 Structure factor, $S(Q)$, for liquid P at several pressures. The letters in parentheses indicate data points in Fig. 1 where the measurements were performed. Energy dispersive X-ray diffraction was performed using white X-rays. To cover a wide wavenumber range, diffraction data were collected at several diffraction angles, from 3.0° to 20.0° . The structure factor $S(Q)$ was obtained using a program developed by Funakoshi³⁰ based on an empirical method proposed by Tsuji¹⁵.

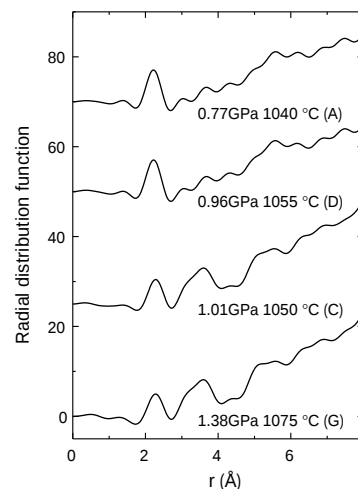


Figure 3 Radial distribution function at several pressures derived from $S(Q)$ for liquid P. The letters in parentheses indicate data points in Fig. 1. r , distance.

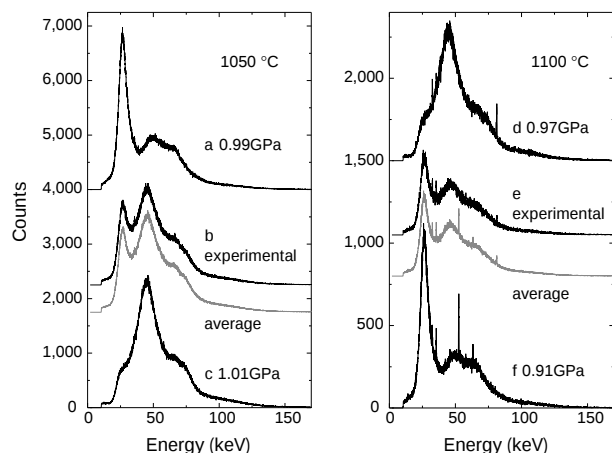


Figure 4 X-ray diffraction patterns. The transformation from the low-pressure form to the high-pressure form (patterns a to c) and the reverse transformation (patterns d to f) are shown. The data were taken at a diffraction angle of 6.0° . The patterns a, c, d and f were measured at data points B, C, E and F in Fig. 1, respectively, and pattern b (pattern e) was measured between data points B and C (data points E and F). The data acquisition times were 300 s for a–c, 120 s for d and f, and 77 s for e. Sharp spikes in patterns d–f are diffraction lines from the BN sample capsule and one unidentified peak. The grey line in pattern b (pattern e) indicates the weighted average of diffraction patterns a and c (patterns d and f).

red P above 350°C (ref. 11). One possible factor is the density. The thermal expansion of molecular liquid P is large¹¹, so perhaps the large separation between molecules at high temperatures may prevent reactions between molecules. The sharpness of the transformation is another mystery. We note that a simulation study of the polymerization of molecular liquid P showed that the polymerization spreads through the sample via “a defect mediated chain reaction”¹⁴. There must be definite pressure–temperature conditions that trigger this chain reaction. The transformation from molecular to polymeric liquid reminds us of the so-called λ transition in liquid sulphur, which is attributed to formation of long-chain polymer components from S_8 ring molecules. However, a recent neutron diffraction study on liquid sulphur has shown that the structural changes accompanying the λ transition are relatively small²⁵. To our knowledge, the present study is the first *in situ* observation of an abrupt structural change in a liquid of a pure substance.

The features of the transformation that we observe in our experiments—the existence of two distinct forms of liquid, a sharp transformation between them, and no detectable intermediate state—strongly support the view that it is a first-order liquid–liquid phase transition. The two forms must have different densities: this is because they have different local atomic structures, and because the transformation pressure coincides with the melting-curve maximum. Unfortunately, the reported melting curve was determined by a limited number of data points, so that the abrupt change of slope was not detected. Precise determination of the melting curve would give useful information.

A liquid–liquid transition has been proposed to occur for several materials, such as H_2O , SiO_2 , GeO_2 , C, Si, Ge, Bi, Se, Te and I (refs 2–5). In addition, coexistence of two liquid phases with the same composition but different densities has been observed in the supercooled melt of $\text{Al}_2\text{O}_3\text{--Y}_2\text{O}_3$ (ref. 26). Among these examples, a first-order transition between low- and high-density liquid water in the supercooled region⁶ has been supported by recent experimental and simulation studies^{7,8}—it is analogous to a transition between low-

and high-density amorphous ice²⁷. We note that there is a rough resemblance between the transitions in supercooled water and in liquid P. Both liquids have tetrahedral units below the transition pressure: tetrahedral pentamer clusters for water⁵ and P_4 molecules for P. In low-density liquid water, these clusters form a hydrogen-bonded network structure. In molecular liquid P, it is certain that the tetrahedral shape of the molecules introduces open space between them^{13,22}. On applying sufficient pressure, these open structures collapse to give new phases with higher density; a polymeric phase in the case of P, and a phase characterized by interpenetrating tetrahedra in the case of water⁵.

The present study is consistent with the general argument that the best candidates to exhibit a liquid–liquid transition are liquids having open molecular coordination environments, especially those with a locally tetrahedral molecular structure². However, the present study also shows that they are not restricted to liquids having extended network structure at low pressures, or to supercooled liquids. The proposed transition in liquid carbon^{9,10} is a transition between thermodynamically stable phases, as in the case of P. It is worth pointing out that black P is remarkably similar to the graphite form of carbon in its crystal structure, anisotropy, and melting behaviour¹. □

Received 25 June; accepted 5 November 1999.

- Young, D. A. *Phase Diagram of the Elements* (Univ. California Press, 1991).
- Poole, P. H., Grande, T., Angell, C. A. & McMillan, P. F. Polymorphic phase transitions in liquids and glasses. *Science* **275**, 322–323 (1997).
- Brazhkin, V. V., Popova, S. V. & Voloshin, R. N. High-pressure transformations in simple melts. *High Pressure Res.* **15**, 267–305 (1997).
- Ponyatovsky, E. G. & Barkalov, O. I. Pressure-induced amorphous phases. *Mater. Sci. Rep.* **8**, 147–191 (1992).
- Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
- Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. Phase behaviour of metastable water. *Nature* **360**, 324–328 (1992).
- Harrington, S., Zhang, R., Poole, P. H., Sciortino, F. & Stanley, H. E. Liquid–liquid phase transition: Evidence from simulations. *Phys. Rev. Lett.* **78**, 2409–2412 (1997).
- Mishima, O. & Stanley, H. E. Decompression-induced melting of ice IV and the liquid–liquid transition in water. *Nature* **392**, 164–168 (1998).
- Togaya, M. Pressure dependence of the melting temperature of graphite and the electrical resistivity of liquid carbon. *Phys. Rev. Lett.* **79**, 2474–2477 (1997).
- Glosli, J. N. & Ree, F. H. Liquid–liquid phase transformation in carbon. *Phys. Rev. Lett.* **82**, 4659–4662 (1999).
- Peck, D. R. in *Mellor's Comprehensive Treatise on Inorganic and Theoretical Chemistry* Vol. VIII, Supp. III, Phosphorus 149–227 (Longman, London, 1971).
- Donohue, J. *The Structure of the Elements* (Wiley & Sons, New York, 1974).
- Clarke, J. H., Dore, J. C., Granada, J. R., Reed, J. & Walford, G. Neutron diffraction studies of liquid phosphorus I. Reactor and pulsed neutron measurements at 50°C . *Mol. Phys.* **42**, 861–874 (1981).
- Hohl, D. & Jones, R. O. Polymerization in liquid phosphorus: Simulation of a phase transition. *Phys. Rev. B* **50**, 17047–17053 (1994).
- Tsuji, K., Yoda, K., Imai, M., Shimomura, O. & Kikegawa, T. Measurements of X-ray diffraction for liquid metals under high pressure. *Rev. Sci. Instrum.* **60**, 2425–2428 (1989).
- Tsuji, K. Structure of liquid metals under high pressure. *J. Non-Cryst. Solids* **117–118**, 27–34 (1990).
- Katayama, Y. *et al.* Density measurements of liquid under high pressure and high temperature. *J. Synchrotron Rad.* **5**, 1023–1025 (1998).
- Katayama, Y., Tsuji, K., Oyanagi, H. & Shimomura, O. Extended X-ray absorption fine structure study on liquid selenium under pressure. *J. Non-Cryst. Solids* **232–234**, 93–98 (1998).
- Akama, Y. *et al.* Melting curve of black phosphorus. *Phys. Lett. A* **122**, 129–131 (1987).
- Kikegawa, T. *et al.* Synchrotron-radiation study of phase transitions in phosphorus at high pressures and temperatures. *J. Appl. Crystallogr.* **20**, 406–410 (1987).
- Rapport, E. Model for melting-curve maxima at high pressure. *J. Chem. Phys.* **46**, 2891–2895 (1967).
- Misawa, M. Structure factor of X_4 tetrahedral molecular liquids: Competition between intramolecular and intermolecular atomic spacings. *J. Chem. Phys.* **93**, 6774–6778 (1990).
- Elliot, S. R., Dore, J. C. & Marseglia, E. The structure of amorphous phosphorus. *J. Phys. C* **8**, 349–353 (1985).
- Hohl, D. & Jones, R. O. Amorphous phosphorus: A cluster-network model. *Phys. Rev. B* **45**, 8995–9005 (1992).
- Winter, R. *et al.* The structural properties of liquid sulphur. *J. Phys.: Condens. Matter* **2**, 8427–8437 (1990).
- Aasland, S. & McMillan, P. F. Density-driven liquid–liquid phase separation in the system $\text{Al}_2\text{O}_3\text{--Y}_2\text{O}_3$. *Nature* **369**, 633–636 (1994).
- Mishima, O., Calvert, L. D. & Whalley, E. An apparently first-order transition between two amorphous phases of ice induced by pressure. *Nature* **314**, 76–78 (1985).
- Zhao, Y., von Dreile, R. B., Weidner, D. J. & Schiferl, D. *P-V-T* data of hexagonal boron nitride, hBN, and determination of pressure and temperature using thermoelastic equation of state of multiphases. *High Pressure Res.* **15**, 369–386 (1997).
- Endo, S., Akahama, Y., Terada, S. & Narita, S. Growth of large single crystals of black phosphorus under high pressure. *Jpn J. Appl. Phys.* **21**, L482–L484 (1982).

30. Funakoshi, K. & Kawamura, K. A novel intensity correction method for energy-dispersive X-ray diffraction using synchrotron radiation: Application to SiO₂ glass at high-pressure. *Acta Crystallogr. A* (submitted).

Acknowledgements

We thank K. Tsuji and Y. Akahama for discussions.

Correspondence and requests for materials should be addressed to Y.K. (e-mail: katayama@spring8.or.jp).

Nanoscale channel lattices with controlled anisotropic wetting

M. Gleiche, L. F. Chi & H. Fuchs

Physikalisches Institut, Westfälische Wilhelms-Universität, D-48149 Münster, Germany

Engineered microscopic surface structures allow local control of physical surface properties such as adhesion, friction and wettability. These properties are related both to molecular interactions and the surface topology^{1,2}—for example, selective adsorption and molecular recognition capabilities³ require controlled anisotropy in the surface properties. Chemistry with extremely small amounts of material has become possible using liquid-guiding channels of sub-micrometre dimensions^{4–6}. Laterally structured surfaces with differing wettabilities may be produced using various techniques, such as microcontact printing^{7–9}, micromachining¹⁰, photolithography^{11,12} and vapour deposition¹³. Another strategy¹⁴ for introducing anisotropic texture is based on the use of the intrinsic material properties of stretched ultrathin polymer coatings. Here we present a fast and simple method to generate extended patterned surfaces with controlled wetting properties on the nanometre scale, without any lithographic processes. The technique utilizes wetting instabilities that occur when monomolecular layers are transferred onto a solid substrate. The modified surfaces can be used as templates for patterning a wide variety of molecules and nanoclusters into approximately parallel channels, with a spatial density of up to 20,000 cm^{−1}. We demonstrate the transport properties of these channels for attolitre quantities of liquid.

Nano-lath lattices consisting of stripes and channels with alternating wettability were produced by using wetting instabilities during the Langmuir–Blodgett film transfer, which has long been used for building up thin organic coatings of monomolecular layers (monolayers) onto solid substrates¹⁵. We used monolayers of L- α -dipalmitoyl-phosphatidylcholine (DPPC) on mica to generate a structured surface with a channel lattice exhibiting a high wettability contrast. This lateral structure can be obtained by rapidly withdrawing a mica substrate (1,000 $\mu\text{m s}^{-1}$) at a low monolayer surface pressure and constant temperature. Under these conditions the film adsorption becomes unstable, leading to periodic interruptions in the molecular deposition. Surface topography inspections with a scanning force microscope (SFM) reveal regular hydrophilic channels of about 200 nm in width, which are separated by hydrophobic stripes of monomolecular height and a latitude of about 800 nm as shown in Fig. 1. These structures extend over macroscopic distances and are only limited by the substrate size. This method is fast and reliable: only one minute is required, for example, to produce a sample of 6 cm in length fully covered with the channel lattice structure shown in Fig. 1 (channel density 10,000 cm^{−1}).

Several explanations have to be taken into account to understand

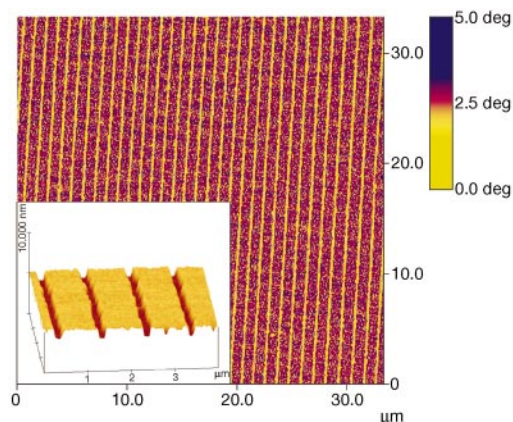


Figure 1 Dynamic scanning force microscope (SFM) image of the structured surface. SFM inspections reveal channels of about 200 nm in width separated by stripes (800 nm wide) of a monomolecular L- α -dipalmitoyl-phosphatidylcholine (DPPC) film, in phase (main figure) and topography (inset, $4 \times 4 \mu\text{m}^2$) imaging. The monolayer prepared on pure water was rapidly transferred (1,000 $\mu\text{m s}^{-1}$) at a lateral surface pressure of 3.0 mN m^{−1} and constant temperature of 22.5 °C. DPPC was purchased from Sigma and used without further purification. A change in temperature will influence the periodicity (lower temperature increases the period), and impurities on the substrate or in the monolayer induce deviations in stripe formation. Deviations in temperature were about ± 0.5 °C leading to variations in the periodicity of 1.0–1.6 μm with channel width of 200–300 nm. Langmuir–Blodgett monolayer preparations (NIMA 6100) and SFM inspections (Digital Instruments, Nanoscope IIIa, Dimension 3000) were carried out with commercial instruments. The film deposition was performed with a commercial d.c.-motor driven Dipper (NIMA).

the observed wetting instability. It is well known that monolayers of an amphiphilic organic molecule spread at the air–water interface can be transferred during a constant vertical upstroke of a solid substrate immersed in the subphase (hydrophilic deposition). The contact line between the water and the solid substrate (three-phase contact line) normally exceeds the planar water surface by a few millimetres as a result of surface tension and defines the static meniscus height. The transfer behaviour of monolayers onto solid substrates depends strongly on the interaction of the amphiphilic molecule with the substrate¹⁶. A strong interaction will result in a rapid adherence of the molecules, causing a reduction in the surface energy¹⁷ and an increase of the contact angle as described by the well known Young's equation. Consequently, the meniscus height will decrease¹⁸. On the other hand, the meniscus height will tend to exceed its equilibrium value due to the continuous substrate motion (upstroke) resulting in an accelerated molecular adsorption on the substrate¹⁹. Thus, the dynamic behaviour of the meniscus height is governed by two counteracting processes, which may lead to an oscillation of the meniscus and result in regions of transfer depletion normal to the dipping direction. A non-uniform stripe formation perpendicular to the transfer direction generated by such a dynamic wetting instability was reported several years ago^{17,19}. A similar stick-jump motion of the meniscus has been found in monolayers of dimethyldioctadecylammonium bromide, resulting in a striped surface, but on a significantly larger length scale (~ 1 mm; ref. 20). Another explanation for the decrease in the contact angle is based on the contribution of the increased line tension due to structural fluctuations of the contact line. It is known from the process of directional solidification that a flat interface between a liquid and a solid becomes unstable while moving through a temperature gradient, resulting in undulated line structures²¹. The additional line energy should increase the contact angle and thus reduce the meniscus height resulting in a deposition-free gap.

In our experiments the oscillation of the meniscus may become periodic, by appropriately adjusting the experimental conditions

such as deposition speed, temperature, and surface pressure. Slow monolayer deposition will result in large distributions of channel distances and widths, whereas an improvement of the regularity is observed with increasing transfer velocity. The deposition speed always exhibits an upper limit where the molecular transfer ratio decreases, although the best quality of the stripe structure (that is, homogeneous periodicity over the whole substrate) is obtained close to this critical value, at which the meniscus oscillation becomes self-synchronized. Nevertheless, a small variation in channel widths and distances as well as small imperfections in the channel boundaries must be taken into account.

The formation of the observed stripe structure can only take place during the molecular transfer of a low-pressure precursor film. The transfer from such an isotropic liquid expanded state (no predefined orientations) is normally not expected to build up close-packed monolayers. In our experiments the transfer ratio was approximately unity, but only 70%–80% of the surface is actually covered. This corresponds well to the difference in the molecular packing density between the condensed phase on the mica surface and the expanded phase on the water surface. The mechanism of stripe formation discussed above differs from that which is responsible for striation in multilamellar liposomes at lower temperature (ripple phase)²² or in simple granular mixtures²³ and in shrinking experiments²⁴. Anisotropic domain growth at the air–water interface can also lead to stripe patterns, which can be understood in terms of a lowered line tension competing with dipole–dipole interactions^{25,26}. There, however, the resulting structure has a random distribution of alignment. The transfer-induced regular

structures reported here differ considerably from stripe phases of monolayers grown at the air–water interface, because of their highly reproducible regularity and predefined orientation parallel to the three-phase contact line.

This structured surface can be used as a template with high adsorption selectivity. The different wetting behaviour of the hydrophilic channels and the hydrophobic stripes can be used to deposit materials along the channels, for example, by either an anisotropic wetting/dewetting process (determined by the surface structure)^{13,27} or by using capillary (capillary filling)^{28,29} and electrostatic forces⁴. The variety of applications is demonstrated in the following three examples.

First, surface-structure-dependent adsorption is obtained by applying a suitable solution carrying the material to deposit. Gold clusters (Au₅₅) stabilized by an organic ligand shell dissolved in 1-phenyloctane were dropped on the structured mica surface and investigated with SFM after the drop has been removed. Exposed to this solvent the DPPC film remains absolutely stable. Channels filled with Au₅₅ clusters are shown (Fig. 2, top). Large cluster aggregates were lined up in the channels. Their height (almost 4 nm) corresponds to a single cluster diameter, but some of them exceed the monolayer height by about one order of magnitude (see spike structure in Fig. 2, bottom). In contrast, the DPPC monolayer is virtually not wetted and only a few cluster aggregates are located on the top of the stripe region, resulting in a nearly perfect selective adsorption.

Second, dye molecules in an aqueous solution fill the channels due to the capillary force, as depicted in Fig. 3. This solution is

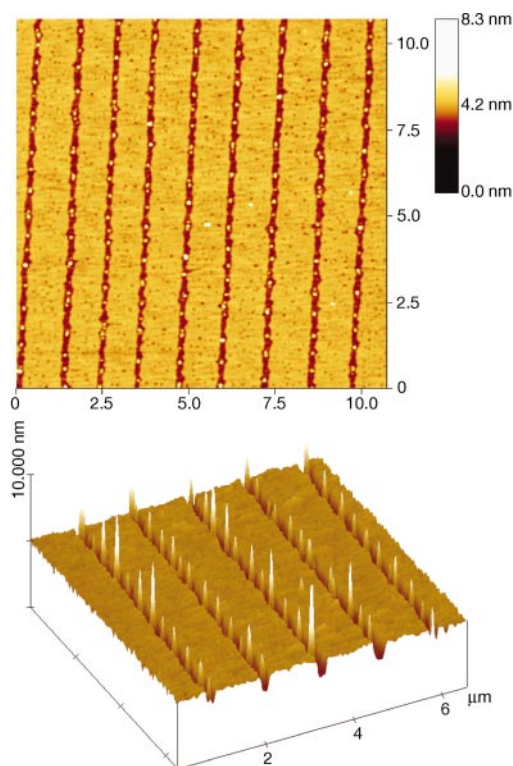


Figure 2 SFM image of liquid-deposited metal cluster. Top: topography of the selectively adsorbed organic-ligand-stabilized metal cluster. Cluster aggregates (bright spots) were aligned along the channels even when scanning the previously wetted area. Bottom: the spikes depicted in the three-dimensionally rendered selected area indicate the large height difference (image size $6.5 \times 6.5 \mu\text{m}^2$).

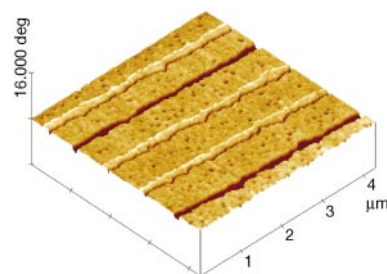


Figure 3 SFM image (phase) of liquid-guiding channels. SFM phase imaging provides a material contrast, which distinguishes between filled and unfilled channels. An aqueous solution carrying dye molecules has been brought into contact with the structured mica surface and channels (bright) were filled up by capillary forces. Two channels shown here (dark) were not filled due to a connection failure caused by the solution. This image has been taken several millimetres away from the wetted area.

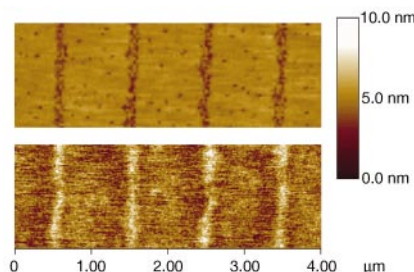


Figure 4 SFM image of gas-deposited FeCl₃ molecules. Top: topographical image of the channel structure filled with FeCl₃ molecules. Bottom: detecting magnetic interactions with magnetic force microscopy on selective adsorbed FeCl₃ molecules in the channels. The measurement was carried out in the 'lift mode' with a scanning height of 70 nm. Topographical cross-talk can thus be excluded. (Lift mode involves scanning at a distance, where topographical contrast is not interfering with weak long-range forces. In addition, the mode retraces the topographical lines as previously measured close to the surface.)

found to cause a partial change in the monolayer morphology resulting in some blocked channels that were not completely filled. Unfilled channels (shown as dark in Fig. 3) can be clearly distinguished from filled ones in the SFM phase image due to the difference in their viscoelasticity³⁰. This material contrast has been pursued over macroscopic distances moving along the 'attolitre' liquid cavities formed by the channels (a single channel that is 1 cm in length can hold up to four attolitres).

Third, a small droplet of FeCl₃ solution was brought onto the structured mica surface and evaporated slowly. The FeCl₃ molecules condensing from the vapour phase were selectively adsorbed in the guiding channels, whereas the monolayer stripes were not coated (Fig. 4, top). The SFM images were taken at positions several centimetres away from the droplet. Channels filled with FeCl₃ molecules provide a contrast in magnetic force microscopy as shown in Fig. 4, bottom.

This method should not be restricted to a specific adsorbate-and-substrate pair and may be extended to other rapidly adsorbing amphiphilic molecules and polymers; however, no other distinct adsorbate-and-substrate pair has yet been investigated. The tiniest channels obtained so far with this method are about 100 nm in width corresponding to a channel density of 20,000 cm⁻¹. Such nano-channel arrays over macroscopic areas may find potential applications as chemical or biochemical reaction cavities with attolitre capacity, in analytical separation techniques, and as textured surfaces.

Received 28 June; accepted November 1999.

- Barthlott, W. & Neinhuis, C. Purity of the sacred lotus, or escape from contaminations in biological surfaces. *Planta* **202**, 1–8 (1997).
- Wagner, T., Neinhuis, C. & Barthlott, W. Wettability and contaminability of insect wings as a function of their surface sculpture. *Acta Zool.* **77**, 213–225 (1996).
- Böltau, M., Walheim, J., Mlynek, J., Krausch, G. & Steiner, U. Surface-induced structure formation of polymer blends on patterned surfaces. *Nature* **391**, 877–879 (1998).
- Gallardo, B. S. *et al.* Electrochemical principles for active control of liquids on submillimeter scales. *Nature* **283**, 57–60 (1999).
- Delamarche, E., Bernard, A., Schmid, H., Michel, B. & Biebuyck, H. Patterned delivery of immunoglobulins to surfaces using microfluidic networks. *Science* **276**, 779–781 (1997).
- Harrison, D. J. *et al.* Micromachining a miniaturized capillary electrophoresis-based chemical analysis system on a chip. *Science* **261**, 895–897 (1993).
- Kumar, A., Biebuyck, H. A. & Whitesides, G. M. Patterning self-assembled monolayers: applications in materials science. *Langmuir* **10**, 1498–1511 (1994).
- Xia, Y. & Whitesides, G. M. Extending microcontact printing as a microlithographic technique. *Langmuir* **13**, 2059–2067 (1997).
- Evans, S. D., Flynn, T. M. & Ulman, A. Self-assembled multilayer formation on predefined templates. *Langmuir* **11**, 3811–3814 (1995).
- Abbott, N. L., Folkers, J. P., Whitesides, G. M. Manipulation of the wettability of surfaces on the 0.1–1-micrometer scale through micromachining and molecular self-assembly. *Science* **257**, 1380–1382 (1992).
- Wang, R. *et al.* Light induced amphiphilic surfaces. *Nature* **388**, 431–432 (1997).
- Calvert, J. M. Lithographic patterning of self-assembled films. *J. Vac. Sci. Technol. B* **11**(6), 2155–2163 (1993).
- Gau, H., Herminghaus, S., Lenz, P. & Lipowsky, R. Liquid morphologies on structured surfaces: from microchannels to microchips. *Science* **283**, 46–49 (1999).
- Wittmann, J. C. & Smith, P. Highly oriented thin films of poly(tetrafluoroethylene) as a substrate for oriented growth of materials. *Nature* **352**, 414–417 (1991).
- Kuhn, H., Möbius, D. & Bücher, H. In *Physical Methods of Chemistry*, Vol. 1 (3b) (eds Weisenberger, A. & Rossiter, B.) 577–702 (Wiley, New York, 1972).
- Petrov, J. G., Kuhn, H. & Möbius, D. Three-phase contact line motion in the deposition of spread monolayers. *J. Colloid Interface Sci.* **37**, 66–75 (1980).
- Riegler, H. & Spratte, K. Structural changes in lipid monolayers during the Langmuir-Blodgett transfer due to substrate / monolayer interactions. *Thin Solid Films* **210/211**, 9–12 (1992).
- Neumann, A. W. Contact angles and their temperature dependence: thermodynamic status, measurement, interpretation and application. *Adv. Colloid Interface Sci.* **59**, 105–191 (1972).
- Spratte, K., Chi, L. F. & Riegler, H. Physisorption instabilities during dynamic Langmuir wetting. *Europhys. Lett.* **25**, 211–217 (1994).
- Erikson, L. G. T., Cleason, P. M., Ohnishi, S. & Hato, M. Stability of dimethyldioctadecylammonium bromide Langmuir-Blodgett films on mica in aqueous salt solutions—implications for surface force measurements. *Thin Solid Films* **300**, 240–255 (1997).
- Langer, J. S. Issues and opportunities in materials research. *Phys. Today* **24–31** (1992).
- Zasadzinski, J. A. N. & Schneider, M. B. Ripple wavelength, amplitude, and configuration in lyotropic liquid crystals as a function of effective headgroup size. *J. Phys. (France)* **48**, 2001–2011 (1987).
- Maske, H. A., Havlin, S., King, P. R. & Stanley, E. Spontaneous stratification in granular mixtures. *Nature* **386**, 379–382 (1997).
- Bowden, N., Brittain, S., Evans, A. G., Hutchinson, J. W. & Whitesides, G. M. Spontaneous formation of ordered structures in thin films of metals supported on an elastomeric polymer. *Nature* **393**, 146–149 (1998).

- Weis, R. M. & McConnell, H. M. Cholesterol stabilizes the crystal-liquid interface in phospholipid monolayers. *J. Phys. Chem.* **89**, 4453–4459 (1985).
- Chunbo, Y. *et al.* Lanthanide ion induced formation of stripes domain structure in phospholipid Langmuir-Blodgett monolayers film observed by atomic force microscopy. *Surf. Sci.* **366**, L729–L734 (1996).
- Biebuyck, H. A. & Whitesides, G. M. Self-organisation of organic liquids on patterned self-assembled monolayers of alkanthiols on gold. *Langmuir* **10**, 2790–2793 (1994).
- Kim, E., Xia, Y. & Whitesides, G. M. Polymer microstructures formed by moulding in capillaries. *Nature* **376**, 581–584 (1995).
- Xia, Y. & Whitesides, G. M. Soft Lithography. *Annu. Rev. Mater. Sci.* **28**, 153–184 (1998).
- Anczykowski, B., Gotsmann, B., Fuchs, H., Cleveland, J. P. & Elings, V. B. How to measure energy dissipation in dynamic mode atomic force microscopy. *Appl. Surf. Sci.* **140**, 376–382 (1999).

Acknowledgements

We thank G. Schmid for providing the Au₅₅ clusters. This work was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to L.F.C. (e-mail:chi@nwz.uni-muenster.de).

DNA computing on surfaces

Qinghua Liu^{*†}, Liman Wang^{*}, Anthony G. Frutos^{*†}, Anne E. Condon^{†‡}, Robert M. Corn^{*} & Lloyd M. Smith^{*}

^{*} Department of Chemistry, [†] Department of Computer Science, University of Wisconsin, Madison, Wisconsin 53706, USA

DNA computing was proposed¹ as a means of solving a class of intractable computational problems in which the computing time can grow exponentially with problem size (the 'NP-complete' or non-deterministic polynomial time complete problems). The principle of the technique has been demonstrated experimentally for a simple example of the hamiltonian path problem² (in this case, finding an airline flight path between several cities, such that each city is visited only once³). DNA computational approaches to the solution of other problems have also been investigated^{4–9}. One technique^{10–13} involves the immobilization and manipulation of combinatorial mixtures of DNA on a support. A set of DNA molecules encoding all candidate solutions to the computational problem of interest is synthesized and attached to the surface. Successive cycles of hybridization operations and exonuclease digestion are used to identify and eliminate those members of the set that are not solutions. Upon completion of all the multi-step cycles, the solution to the computational problem is identified using a polymerase chain reaction to amplify the remaining molecules, which are then hybridized to an addressed array. The advantages of this approach are its scalability and potential to be automated (the use of solid-phase formats simplifies the complex repetitive chemical processes, as has been demonstrated in DNA and protein synthesis¹⁴). Here we report the use of this method to solve a NP-complete problem. We consider a small example of the satisfiability problem (SAT)², in which the values of a set of boolean variables satisfying certain logical constraints are determined.

Our overall strategy for DNA computing on surfaces has been described in detail previously¹⁰, and consists of six main steps shown diagrammatically in Fig. 1. Each step in the process must be reasonably efficient for the overall process to succeed; the development of these steps has been the subject of previous work^{10–12}, and may be briefly summarized as follows. Oligonucleotides are synthesized individually, and pooled or arrayed on surfaces as needed. The 5' thiol-modified oligonucleotides (S_xs) are attached to the surface in an unaddressed fashion; compared to addressed oligonucleotide

[†] Present addresses: Gen-Probe, Inc., San Diego, California 92121, USA (Q.L.); Corning Inc., Corning, New York 14831, USA (A.G.F.); Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (A.E.C.).

arrays^{15,16}, this allows a greater number of different sequences to be deposited per unit area, but necessitates the use of a final 'readout' step to determine the identity of the remaining oligonucleotides once the computation has been performed¹⁰. Each DNA 'word'¹¹ consists of 16 nucleotides with the structure 5'-FFFFvvvvvvFFFF-3'; the internal eight nucleotides ('v') are variables and encode the information (Table 1); the eight bases labelled 'F' are fixed 'word labels' that direct DNA hybridization of 16 nucleotide complements to that word in the 'mark' operation. The prototype DNA computer described here uses a single word. SAT computation is done through repeated cycles of 'mark', 'destroy' and 'unmark' operations¹⁰ (Fig. 1). Finally, determination of the 'answer' to the computational problem of interest is accomplished in a 'readout' operation (see Methods). We note that other investigators have also reported surface-based approaches to DNA computing; indeed, Adleman's first report of DNA computing utilized support chemistry to separate DNA molecules¹; and more recently, Morimoto *et al.*¹⁷ and Yoshida and Suyama¹⁸ have described solid-phase procedures for DNA computing.

The SAT problem is an NP-complete problem in boolean logic². An instance of the SAT problem consists of a set of boolean logic variables separated by the logical OR operation (denoted by " $\vee$ "; $u \vee v = 0$ if and only if $u = v = 0$) within clauses, and with the clauses separated by the logical AND operation (denoted by " $\wedge$ "; $u \wedge v = 1$ if and only if $u = v = 1$). The problem is to find whether there are values for the variables that simultaneously satisfy each clause in a given instance of the problem². The specific SAT problem solved here is:

$$(w \vee x \vee y) \wedge (w \vee \bar{y} \vee z) \wedge (\bar{x} \vee y) \wedge (\bar{w} \vee \bar{y})$$

and employs four variables w, x, y and z ($\bar{w}, \bar{x}, \bar{y}$ and $\bar{z}$ denote the negation of the variables w, x, y and z ; thus $\bar{w} = 0$ if and only if $w = 1$,

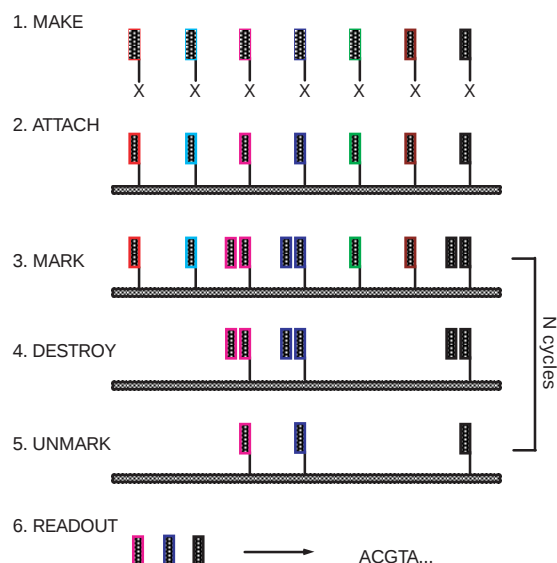


Figure 1 Overview of the surface-based approach to DNA computations. A combinatorial set of single-stranded DNA molecules representing all possible solutions to a given computational problem is synthesized ('make') and immobilized ('attach') on a surface via a reactive functional group X. In each of N successive cycles of the DNA computation, subsets of the surface-bound combinatorial mixture are tagged by hybridization to their complements in a 'mark' operation, rendering them double-stranded. After the 'mark' operation, an enzyme (for example, *Escherichia coli* exonuclease I) is added which destroys surface-bound oligonucleotides present in an unhybridized single-stranded form ('destroy'). The surface is then regenerated by removing all hybridized complements in an 'unmark' operation. Repetitive cycles of 'mark', 'destroy' and 'unmark' operations remove from the surface all strands which do not satisfy the problem. At the end of N cycles, only those strands which are solutions to the problem remain. Their identities are determined in a 'readout' operation by PCR followed by hybridization to an addressed array.

and $\bar{w} = 1$ if and only if $w = 0$). Each of the four variables can be either true (1) or false (0) and thus there are total of 2^4 or 16 candidate solutions. All methods for solving the SAT problem (on any realizable computing device) that have been analysed require a number of steps that grows exponentially with the number of variables. However, algorithms have been developed for conventional computing which substantially decrease the difficulty of the problem compared to a 'brute-force' search through all possible solutions. For example, the best known algorithm for the 3-SAT problem, in which the number of variables per clause is at most 3, has expected running time bounded by a polynomial time 1.33^n for large n (ref. 19), a very substantial improvement compared to the 2^n running time characteristic of the 'brute-force' search. It has recently been shown that such algorithms may also be implemented in DNA computing approaches, including specifically the 3-SAT problem addressed here^{6,18}. We show here how 3-SAT may be solved using immobilized DNA molecules and a simple 'brute-force' search strategy; the more complex operations required to execute more sophisticated algorithms are under development.

The surface-bound oligonucleotide sequences utilized were of the form 5'-HS-C₆-T₁₅GCTTvvvvvvvTTCG-3' (S_x s). The T₁₅ sequence serves as a 'spacer' group to separate the hybridizing sequence from the support²⁰, the GCTT and TTCG sequences are the 'word label' used to target hybridization to a particular word¹¹, and the v₈ sequence is used to encode information. Table 1 shows the 16 octanucleotide sequences used, which were chosen from a previously described set of 108 possible sequences¹¹, along with the encoding scheme utilized to represent the possible values of the SAT variables w, x, y and z . These surface-bound oligonucleotides, representing all candidate solutions, were synthesized, pooled, and attached to a maleimide-functionalized gold surface in an unaddressed format^{21,22}. Each clause of the SAT problem requires one cycle of 'mark', 'destroy' and 'unmark' (Fig. 1), and thus four cycles were employed to solve the example SAT problem above. The goal of the first computational cycle is to destroy all DNA molecules which do not satisfy the first clause ($w \vee x \vee y$). This is achieved by hybridizing to the surface those oligonucleotides that are complementary to the molecules which do satisfy the clause, and then destroying the remaining (unmarked) single-stranded molecules. Only two sequences do not satisfy this clause: namely, those for which w, x and y are set to zero (S_0 [0000] and S_1 [0001], see Table 1). Thus in cycle 1 the complements (C_x s) of the 14 other oligonucleotides ($w = 1$ ($C_8, C_9, C_{10}, C_{11}, C_{12}, C_{13}, C_{14}, C_{15}$); $x = 1$ ($C_4, C_5, C_6, C_7, C_{12}, C_{13}, C_{14}, C_{15}$); $y = 1$ ($C_2, C_3, C_6, C_7, C_{10}, C_{11}, C_{14}, C_{15}$)) were combined and hybridized (in the 'mark' operation) to the surface; after washing, the surface was exposed to *Escherichia coli* exonuclease I to destroy the unhybridized, single-stranded molecules S_0 and S_1 (in the 'destroy' operation). The surface was regenerated

Table 1 Variable sequences and encoding scheme

Strand	Variable sequence	wxyz
S_0	CAACCCAA	0000
S_1	TCTCAGAG	0001
S_2	GAAGGCAT	0010
S_3	AGGAATGC	0011
S_4	ATCGAGCT	0100
S_5	TTGGACCA	0101
S_6	ACCATTGG	0110
S_7	GTTGGGTT	0111
S_8	CCAAGTTG	1000
S_9	CAGTTGAC	1001
S_{10}	TGTTTGG	1010
S_{11}	GATCCGAT	1011
S_{12}	ATATCGCG	1100
S_{13}	GGTTCAC	1101
S_{14}	AACCTGGT	1110
S_{15}	ACTGTGCA	1111

The 16 strands shown encode 4 bits (2^4) of information (4 variables: w, x, y, z). S_x is the 16-nucleotide DNA sequence 5'-GCTTvvvTTCG-3', where the v₈ sequence is shown as the "Variable sequence" above; C_x is the 16-nucleotide complement of the S_x sequence.

by the 'unmark' operation to return the remaining surface-bound oligonucleotides S_2 – S_{15} to single-stranded form. This process was repeated three more times for the remaining three clauses, to yield a surface containing only the solutions to the SAT problem.

In Fig. 2 we show in detail the logic of the DNA computation in each cycle, leading at the end to four types of DNA molecules remaining on the surface. It may be noted that since each clause is tested independently, in some cases complements are added corresponding to surface-bound oligonucleotides that are no longer on the surface, having been destroyed in a previous cycle. Although this entails some redundancy, the independence of clause-testing obtained in this manner increases the computational versatility and power. In the particular SAT problem solved here, each strand is destroyed only once (that is, is targeted in only one of the four computational cycles). The identity of those molecules that correspond to the solutions was determined at the end of the computation by polymerase chain reaction (PCR) amplification and hybridization to an addressed array. The results are shown in Fig. 3, presented both as a fluorescence image and in histogram form.

The four spots with high fluorescence intensity in Fig. 3 correspond to the four expected solutions to the computational problem posed. The DNA sequences identified in the 'readout' step via addressed array hybridization were: S_3 , S_7 , S_8 and S_9 (Fig. 3). Their variable sequences are AGGAATGC, GTTGGGT, CCAAGTTG,

and CAGTTGAC (Table 1), corresponding to the truth assignments of ($w = 0, x = 0, y = 1$, and $z = 1$), ($w = 0, x = 1, y = 1$, and $z = 1$), ($w = 1, x = 0, y = 0$, and $z = 0$), and ($w = 1, x = 0, y = 0$, and $z = 1$). For example, the assignment ($w = 0, x = 0, y = 1$, and $z = 1$) satisfies the SAT problem solved here in that $y = 1$ satisfies the first clause ($w \vee x \vee y$), $z = 1$ satisfies the second clause ($w \vee \bar{y} \vee z$), either $x = 0$ or $y = 1$ makes the third clause ($\bar{x} \vee y$) true, and $w = 0$ satisfies the last clause ($\bar{w} \vee \bar{y}$). The signal intensities for the spots corresponding to the correct solutions were 10 to 777 times greater than those corresponding to incorrect solutions, making discrimination between correct and incorrect solutions to the problem straightforward. We now consider the issues which need to be addressed for scaling of the approach to substantially larger problems.

Each step in the above-described DNA computing process is chemically complex and subject to error. For example, in the 'mark' operation, the nature of chemical reaction kinetics ensures that the hybridization-driven formation of duplex structures on the surface will be incomplete, and hence some surface-immobilized DNA molecules encoding correct answers to the clause in question will not form duplexes; these molecules will then be subject to unwanted destruction during the 'destroy' (exonuclease digestion) step, degrading signal during the computational process. Conversely, mismatch hybridization to incorrect sequences present on the surface may protect them inappropriately from destruction, leading to background signal corresponding to incorrect solutions to the problem.

In addition, the destruction of single-stranded DNA molecules on the surface by *E. coli* exonuclease I is not perfect. Approximately 94% of unmarked single-stranded immobilized DNA molecules can be removed by the exonuclease digestion reaction¹¹. This means that about 6% of the unwanted (incorrect) DNA molecules will remain on the surface after a 'destroy' operation, giving rise to false positive

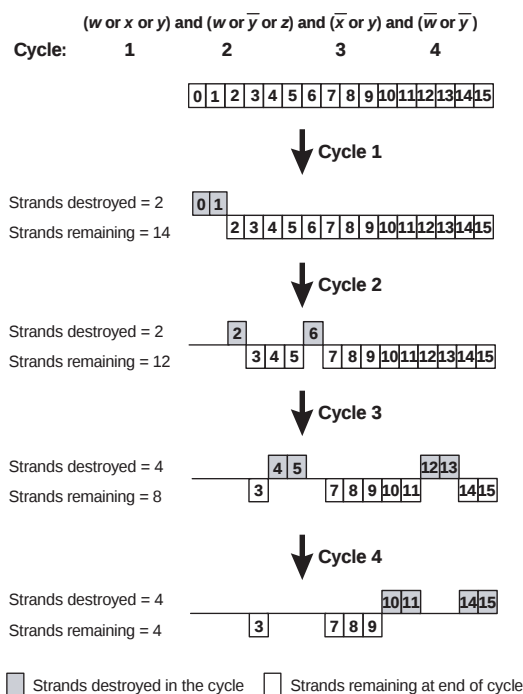


Figure 2 Four cycles of SAT computation. For each cycle of the computation, the number of strands destroyed in that cycle of the computation and the number of strands remaining on the surface at the end of the cycle are indicated on the left. The identities of the destroyed and remaining strands are shown in the progression of the computation in schematic form: in each cycle of computation one of the four clauses is tested by applying the 'mark' operation to only the oligonucleotides that satisfy that clause. For example, in cycle 1, applying the 'mark' operation to S_2 – S_{15} (leaving S_0 and S_1 unmarked) protects all the surface-bound oligonucleotides for which w or x or y has a value of 1. A similar procedure is applied in each of the three subsequent cycles for clauses 2–4, respectively. In each cycle, strands indicated in the shaded boxes are destroyed using the single-strand specific *E. coli* exonuclease I, while strands indicated in the non-shaded boxes are protected by their complements. The strands remaining in the non-shaded boxes at the end of the process are the solutions to the computational problem.

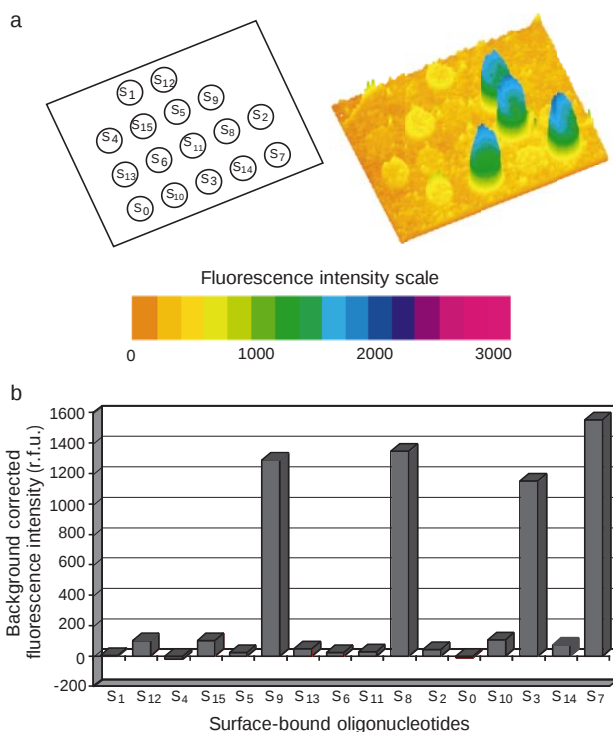


Figure 3 Three-dimensional plot and histogram of the fluorescence intensities on a 16 element addressed array used for 'readout'. **a**, Fluorescence profile (right) with the surface-bound oligonucleotide locations (left). **b**, The fluorescence intensity histogram. The three-dimensional plot in **a** was generated using NIH Image version 1.61 software (National Institutes of Health, Bethesda, Maryland; <http://rsb.info.nih.gov/ni-image/download.html>). r.f.u., relative fluorescence units.

signals. A more fundamental issue with the 'destroy' operation presented here is that it is incompatible with a multiple-word encoding strategy; work is in progress to develop an alternative 'destroy' operation for this purpose.

These errors in the 'mark' and 'destroy' operations also affect the 'readout' process. We have found it to be very difficult to uniformly amplify a population of DNA molecules, even ones as tightly constrained as the set employed here for DNA computing²³ (data not shown). Small amounts of unwanted oligonucleotides, if amplified preferentially compared to desired species present initially at higher concentration, can yield aberrant results. Although the GC content was held constant for the set of DNA molecules used in this work, variations in the position of GC pairs within the sequence can affect the formation of secondary structures that influence the efficiency of PCR amplification. Betaine is known to preferentially destabilize GC base pairs²⁴. The use of betaine here resulted in a more uniform amplification of the mixture of DNA templates²⁵. Nonetheless, substantial optimization and experimentation with the PCR conditions were required to obtain the results shown here, and it is likely that this would become increasingly difficult with larger problems. This is fundamentally a reflection of the intrinsically nonlinear nature of the PCR amplification process²⁶; an alternative method of amplification which is linear in nature is under investigation²⁷.

This intrinsically error-prone nature of the DNA computing process, however, does not rule out its operational practicality. The fundamental operations of conventional computing are also prone to error at the device level, and there is a rich body of knowledge on algorithmic solutions to these issues which could be adapted to the DNA computing process²⁸. Methods for handling errors in DNA computations have been reported^{29,30}. Experimental determination of error rates in surface-based DNA computing could provide the information necessary to develop error models and algorithmic methods to control errors in computations.

The surface approach described here was designed to permit scale-up to larger problems. For example, the use of six tandem words on the surface (for a total oligomer length of $16 \times 6 = 96$ nucleotides, within the range of current oligonucleotide synthesis capabilities), and 64 different octanucleotide sequences¹¹ to encode information in each of the words (6 bits per word), corresponds to a search space containing approximately 6.9×10^{10} candidate solutions^(2³⁶); that is, a 36-bit DNA computer).

The number of 'mark' and 'destroy' operations required to solve the problem grows polynomially with the number of variables¹³, in contrast to the search space which grows exponentially. Thus a 36-bit DNA computer would not look a great deal different from the 4-bit computer described here. The biggest practical issue (apart from those mentioned above) is the synthesis and handling of large numbers of oligonucleotides (1,536 oligonucleotides would be needed to carry out the 'mark' and 'readout' operations for a 6-word, 6 bits per word, DNA computer of this type; see Methods). Instrumentation to synthesize and handle such large numbers of oligonucleotides is available (ABI 3948 Nucleic Acid Synthesis and Purification System (Perkin Elmer); also see refs 15 and 16). Specialized problems might require far fewer oligonucleotides than this (see Methods).

Methods

PCR amplification for 'readout'.

56-mer PCR templates were synthesized with the following structure:

5'-tatttttgagcagtggtccCGAAvvvvvvvAAGCtagtatctacaaggtcag-3'

where the lower-case sequences are the two primers, the upper-case sequences are fixed word labels, and the eight bases labelled 'v' are complements of the variable regions shown in Table 1.

For PCR amplification and addressed array hybridization 'readout', the template(s) used in PCR amplification were collected from the computation chip (that is, the non-

addressed array surface which had undergone repetitive computation operations) as follows: upon completion of each computation cycle (including 'mark', 'destroy', 'unmark'), the coverslip (computation chip) was exposed to a mixture of all 16 complementary DNA strands (each containing two 20-mer primer sequences as indicated above) for 1.5 h for hybridization. The coverslip was then washed; all hybridized complements were melted off and collected by assembling the coverslip in a GeneAmp *In Situ* PCR System 1000 (Perkin Elmer) containing 100 µl of de-ionized water, and heating at 95 °C for 10 min. The collected template(s) were then amplified using biotinylated and fluorescently tagged PCR primers as follows:

5'-biotin-ctgaatctttagatagcta-3' 5'-FAM-tatttttgagcagtggtcc-3'

The amplification reaction (100 µl) contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 3 mM MgCl₂, 0.2 mM each dATP, dCTP, dTTP and dGTP, 1 µM of each primer, 1 M betaine, and 5 units of AmpliTaq DNA polymerase (Perkin Elmer). In addition, 10 fmol of a 56-mer oligonucleotide identical in structure to that shown above, but containing in the variable region the non-complementary 8-mer sequence GAGACTCT, was included in the reactions. The function of this was to help equalize the amount of material that could be amplified in the various PCR reactions performed. This was found to reduce artefacts due to amplification of trace amounts of templates present in the reactions, a source of background in the 'readout' operation. PCR was performed in an PTC-200 Peltier Thermal Cycler (MJ Research, Inc.) using 25 cycles. The PCR mixture was first held at 95 °C for 1 min, followed by cycles of 95 °C for 30 s, 40 °C for 20 s and 60 °C for 1 min.

Hybridization to addressed array for 'readout'.

The DNA addressed array was prepared by methods similar to those described in ref. 31. Briefly, a gold-coated coverslip was immersed in a 0.5 mM ethanolic solution of *n*-octadecylmercaptan overnight, followed by photopatterning of the surface using an aluminium mask consisting of 1.5-mm-diameter holes with 3 mm centre-to-centre spacing. The coverslip was then immersed in 1 mM ethanolic 11-mercaptopundecanoic acid for 1 h, followed by electrostatic adsorption of a poly-L-lysine monolayer, which was maleimide-activated as described previously²¹.

PCR amplification of the products of the DNA computing process yields double-stranded 56-mers with one strand biotinylated, and the other strand fluorescein-labelled. Before addressed array readout, these double-stranded PCR products were strand-separated using streptavidin-coupled magnetic beads (Dynabeads M-280, Dynal Prod. No. 112.05, 112.06 Protocol, 1997, Dynal Inc.). After strand separation, the fluorescently labelled PCR products were resuspended in hybridization buffer (HB; 2XSSPE/0.2%SDS), applied to the 16-spot addressed array and allowed to hybridize for 1.5 h at room temperature. The addressed array coverslip was washed twice in HB, 5 min per wash, followed by immersion in a beaker of HB at 37 °C for 20 min. The coverslip was then scanned using a Molecular Dynamics FluorImager 575.

Number of oligonucleotides required for scale-up.

The oligonucleotides employed for a general purpose multiple word DNA computer of the type employed here are of two types: (1) multiple word combinatorial mixtures, thiol-modified to permit surface attachment when preparing the unaddressed surface on which the computation is performed, and (2) single word individual oligonucleotides, either thiol-modified for surface attachment when preparing addressed 'readout' arrays (*S_s*), unmodified for use in the 'mark' operation (*C_s*), or with PCR-primer sequences appended for use in the 'readout' operation (*C_s* with appended PCR-primer sequences). Thus a total of four sets of oligonucleotides must be synthesized. Although (as described in the text) the number of distinct oligonucleotides present in the combinatorial mixtures can be very large, a much smaller set of distinct oligonucleotide synthesis procedures is required to prepare them using split-and-pool methodology^{23,33}; this is given by the product of the number of different oligonucleotide sequences per word and the number of words. In the 36-bit example provided above, this corresponds to $64 \times 6 = 384$ oligonucleotide syntheses. The number of individual oligonucleotides required for the three sets of *S_s* and *C_s* is given by the same product, for a total of $384 \times 4 = 1,536$; when solving a 3-SAT, in any given cycle (corresponding to a single clause) the number of *C_s* required for the 'mark' operation is at most $(64 / 2) \times 3 = 96$.

Received 7 July; accepted 20 October 1999.

- Adleman, L. M. Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994).
- Garey, M. R. & Johnson, D. S. *Computers and Intractability: a Guide to the Theory of NP-completeness* (Freeman, New York, 1979).
- Adleman, L. M. Computing with DNA. *Sci. Am.* **279**, 54–61 (1998).
- Lipton, R. J. DNA solution of hard computational problems. *Science* **268**, 542–545 (1995).
- Guarnieri, F., Fliiss, M. & Bancroft, C. Making DNA add. *Science* **273**, 220–223 (1996).
- Ogihara, M. *Breadth First Search 3SAT Algorithms for DNA Computers* (Technical report TR 629, Department of Computer Science, University of Rochester, Rochester, New York, 1996).
- Ouyang, Q., Kaplan, P. D., Liu, S. & Libchaber, A. DNA solution of the maximal clique problem. *Science* **278**, 446–449 (1997).
- Seeman, N. C. *et al.* New motifs in DNA nanotechnology. *Nanotechnology* **9**, 257–273 (1998).
- Winfree, E., Liu, F. R., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).
- Smith, L. M. *et al.* A surface-based approach to DNA computation. *J. Comput. Biol.* **5**, 255–267 (1998).
- Frutos, A. G. *et al.* Demonstration of a word design strategy for DNA computing on surfaces. *Nucl. Acid. Res.* **25**, 4748–4757 (1997).
- Frutos, A. G., Smith, L. M. & Corn, R. M. Enzymatic ligation reactions of DNA "words" on surfaces for DNA computing. *J. Am. Chem. Soc.* **120**, 10277–10282 (1998).

13. Cai, W. *et al.* in *Proc. 1st Annu. Int. Conf. on Computational Molecular Biology (RECOMB97)* 67–74 (Association for Computing Machinery, New York, 1997).
14. Smith, L. M. Automated synthesis and sequence analysis of biological macromolecules. *Anal. Chem.* **60**, 381A–390A (1988).
15. Fodor, S. P. A. *et al.* Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).
16. Chee, M. *et al.* Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614 (1996).
17. Morimoto, N., Arita, M. & Suyama, A. in *Proc. DIMACS: DNA Based Computers (III)* (eds Rubin, H. & Wood, D. H.) 83–92 (American Mathematical Society, Providence, 1997).
18. Yoshida, H. & Suyama, A. in *Preliminary Proc. DIMACS: DNA Based Computers (V)* 9–20 (American Mathematical Society, Providence, 1999).
19. Schöning, U. in *Proc. 40th Annu. IEEE Conf. of Foundations of Computer Science (FOCS)* 410–414 (IEEE Computer Society, Los Alamitos, California, 1999).
20. Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucl. Acid. Res.* **22**, 5456–5465 (1994).
21. Jordan, C. E., Frutos, A. G., Thiel, A. J. & Corn, R. M. Surface plasmon resonance imaging measurements of DNA hybridization adsorption and streptavidin/DNA multilayer formation at chemically modified gold surfaces. *Anal. Chem.* **69**, 4939–4947 (1997).
22. Bain, C. D. *et al.* Formation of monolayer films by the spontaneous assembly of organic thiols from solution onto gold. *J. Am. Chem. Soc.* **111**, 321–335 (1989).
23. Baskaran, N. *et al.* Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. *Genome Res.* **6**, 633–638 (1996).
24. Rees, W. A., Yager, T. D., Korte, J. & von Hippel, P. H. Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* **32**, 137–144 (1993).
25. Voss, K. O., Pieter Roos, K., Nonay, R. L. & Dovichi, N. J. Combating PCR bias in bisulfate-based cytosine methylation analysis. Betaine-modified cytosine deamination PCR. *Anal. Chem.* **70**, 3818–3823 (1998).
26. Farrell, R. E. DNA amplification. *Immunol. Invest.* **26**, 3–7 (1997).
27. Lyamichiev, V. *et al.* Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnol.* **17**, 292–296 (1999).
28. Pippenger, N. Developments in the synthesis of reliable organisms from unreliable components. *Proc. Symp. Pure Math.* **50**, 311–324 (1990).
29. Boneh, D., Dunworth, C., Lipton, R. J. & Sgall, J. DNA Based Computers II (eds Landweber, L. F. & Baum, E. B.) 163–170 (DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 44, American Mathematical Society, Providence, 1999).
30. Karp, R. M., Kenyon, C. & Waarts, O. Error-resilient DNA computation. *Random Struct. Algor.* **15**, 450–466 (1999).
31. Gillmor, S. D., Thiel, A. J., Smith, L. M. & Lagally, M. G. Hydrophilic/hydrophobic patterned surfaces as templates for DNA arrays. *Langmuir* (submitted).
32. Gallop, M. A., Barrett, R. W., Dower, W. J., Fodor, S. P. A. & Gordon, E. M. Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J. Med. Chem.* **37**, 1233–1251 (1994).
33. Gordon, E. M., Barrett, R. W., Dower, W. J., Fodor, S. P. A. & Gallop, M. A. Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. *J. Med. Chem.* **37**, 1385–1401 (1994).

Acknowledgements

We thank S. Gillmor and J. Brockman for help with the preparation of the photopatterned read-out arrays, and M. Lagally for discussions. This work was supported by the Defense Advanced Research Projects Agency (DARPA) and the National Science Foundation.

Correspondence and requests for materials should be addressed to L.M.S.
(e-mail: smith@chem.wisc.edu)

Evidence for enhanced mixing over rough topography in the abyssal ocean

J. R. Ledwell, E. T. Montgomery, K. L. Polzin, L. C. St. Laurent, R. W. Schmitt & J. M. Toole

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

The overturning circulation of the ocean plays an important role in modulating the Earth's climate. But whereas the mechanisms for the vertical transport of water into the deep ocean—deep water formation at high latitudes—and horizontal transport in ocean currents have been largely identified, it is not clear how the

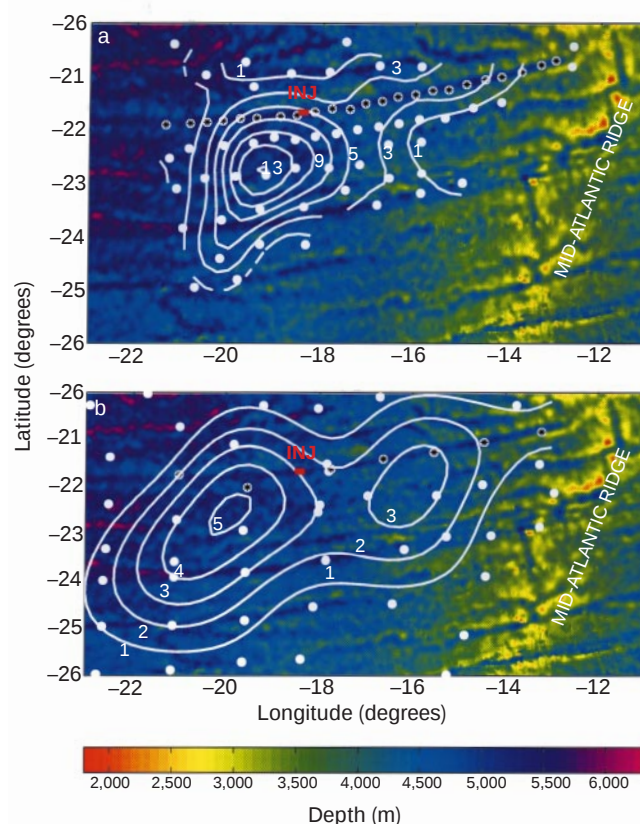


Figure 1 Tracer distribution. **a**, 14 months after release; **b**, 26 months after release. The red bars labelled 'INJ' mark the release site of the tracer. The contours denote the column integral of SF_6 (in nmol m^{-2}) and colours denote bottom depth. The tracer mapping procedure did not take the bathymetry into account, and hence the meridional distribution in the east in **a** appears broader than the valleys where the stations are located and which actually hold most of the tracer. The bathymetry is from Smith and Sandwell¹⁶. The stations are shown as white dots; those with stars are used for the sections in Fig. 2. Southerly latitude and westerly longitude are shown negative.

compensating vertical transport of water from the depths to the surface is accomplished. Turbulent mixing across surfaces of constant density is the only viable mechanism for reducing the density of the water and enabling it to rise. However, measurements of the internal wave field, the main source of energy for mixing, and of turbulent dissipation rates, have typically implied diffusivities across surfaces of equal density of only $\sim 0.1 \text{ cm}^2 \text{ s}^{-1}$, too small to account for the return flow. Here we report measurements of tracer dispersion and turbulent energy dissipation in the Brazil basin that reveal diffusivities of $2\text{--}4 \text{ cm}^2 \text{ s}^{-1}$ at a depth of 500 m above abyssal hills on the flank of the Mid-Atlantic Ridge, and approximately $10 \text{ cm}^2 \text{ s}^{-1}$ nearer the bottom. This amount of mixing, probably driven by breaking internal waves that are generated by tidal currents flowing over the rough bathymetry, may be large enough to close the buoyancy budget for the Brazil basin and suggests a mechanism for closing the global overturning circulation.

Our study was conducted in the abyssal Brazil basin where deep upwelling can be inferred from measurements of net inflow of dense water^{4,5}. In 1996 we surveyed turbulent microstructure and internal wave fine-structure across the basin, and released 110 kg of sulphur hexafluoride above one of the zonal valleys on the western flank of the Mid-Atlantic Ridge⁶ (Fig. 1). The microstructure data showed diapycnal mixing to be very low over the smooth parts of the Brazil

basin in the west, but greatly enhanced throughout much of the water column in the eastern part of the basin, with a distinct increase toward the bottom. The initial dispersion of the tracer during the first two weeks, and the microstructure measurements in its immediate vicinity, indicated diffusivities of about $0.5 \text{ cm}^2 \text{ s}^{-1}$ at the isopycnal surface of the tracer release ($45.9408 \text{ kg m}^{-3}$, referenced to 4,000 dbar). This surface was about 4,000 m deep, approximately 1,000 m above the valley floor and 500 m above the hilltops lying 20–30 km to the north and south of the release site. The microstructure data suggested still larger mixing rates to the east and closer to the bottom⁶.

A second survey of the tracer and microstructure in 1997, 14 months after the release of the tracer, found about 95% of the tracer in a patch centred southwest of the release site (Fig. 1a). The remaining 5% was located as far as 500 km east of the release site, deep in the valley below the release, and as far as 300 km east in another zonal valley lying 200 km to the south. Tracer to the west peaked near the target density surface, while tracer to the east in the valleys was concentrated nearer the bottom (Fig. 2a). A third tracer survey in 1998, 26 months after its release, found about 70% of the tracer in the patch to the west, which had spread laterally and vertically, while the patch in the east had grown to include about 30% of the tracer and had dispersed dramatically across isopycnals (Figs 1b and 2b).

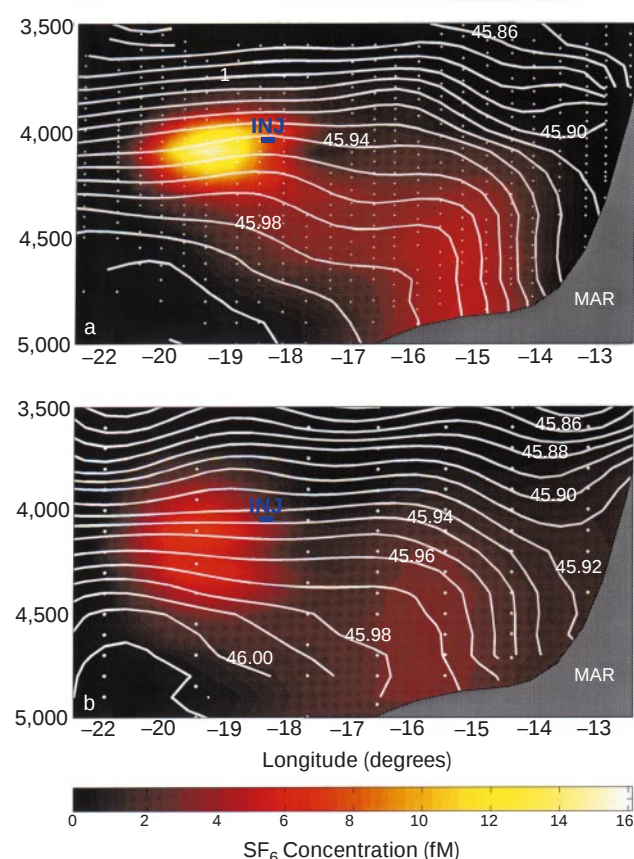


Figure 2 Sections of tracer concentration and potential density from the valley where the tracer was released. **a**, 14 months after release; **b**, 26 months after release. The dots show the tracer sample locations, and the blue bar labelled 'INJ' marks the release site of the tracer. The sea floor (labelled MAR for Mid-Atlantic Ridge) is a polynomial fit to the soundings from the 14-month survey. The valley is enclosed by ridges to the north and south whose depths are indicated roughly where the white density contours (in kg m^{-3}) bend sharply down. The colours show the trace concentration: $1 \text{ fM} = 10^{-15} \text{ M}$.

The inventory of tracer as a function of potential density has been estimated from all of the 1997 data and transformed into a profile of concentration versus height z above the target density surface using the mean density profile west of 18° W (Fig. 3). The diapycnal diffusivity $K(z)$ was estimated by applying a one-dimensional model for the tracer evolution from the initial profile in 1996 to this 1997 profile (see Methods). Least-squares fits, shown in Fig. 3, were obtained for $K = 3.0 \text{ cm}^2 \text{ s}^{-1}$ at $z = 0$ ($\sim 4,000 \text{ m}$ depth), increasing to $8 \text{ cm}^2 \text{ s}^{-1}$ at $z = -500 \text{ m}$ ($\sim 4,500 \text{ m}$ depth). To obtain this fit it was necessary to include a velocity of the tracer relative to the water of $-1.0 \times 10^{-4} \text{ cm s}^{-1}$. We can only suggest that tracer was transported on sinking particles, although the rate is at least a factor of 10 greater than that at a depth of 300 m in the subtropical North Atlantic⁷. It remains to be tested whether particulate effects are enhanced at high pressure and low temperature. This sinking continued at approximately the same rate between 1997 and 1998.

The three-dimensional tracer distributions of Figs 1 and 2 strongly suggest that diapycnal mixing on isopycnals increases toward the east as density surfaces approach the bottom. The one-dimensional model cannot account for this spatial variation, and so inferences from it are only approximate. In particular, because tracer found in the east is concentrated near the bottom, enhanced diffusion there tends to mix tracer from denser isopycnals back toward the target density surface; the model may thus underestimate the diffusivity on the deeper isopycnal surfaces. The estimate of K at the target density surface seems to be accurate to within $\pm 1 \text{ cm}^2 \text{ s}^{-1}$, although the depth dependence of K is not well constrained with a one-dimensional model. Because only 5% of the tracer was in the east in 1997, the one-dimensional model seems justified. In 1998, 30% of the tracer was in the east and the transport back to lighter isopycnals seemed so pronounced that the one-dimensional approach has not been taken.

The diffusivity inferred from measurements of the dissipation of turbulent kinetic energy about the target surface in the region of the tracer patch was $2.3 \pm 1 \text{ cm}^2 \text{ s}^{-1}$, similar to the tracer-derived value when adjustments for spatial and temporal biases in the sampling are taken into account (see Methods). Microstructure profiles clearly show an increase of diffusivity with decreasing height above the bottom (Fig. 4). A spatial bias toward low average diffusivity resulted from the location of most of the stations over valleys where the target isopycnal lay about 1,000 m above the local

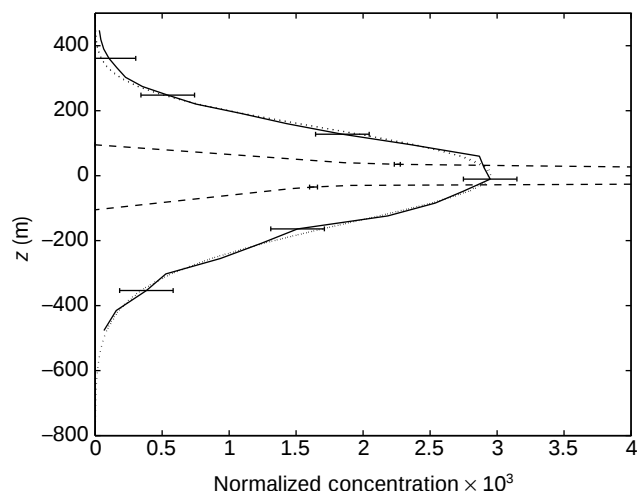


Figure 3 Mean vertical tracer profiles (with uncertainties) and model fit. The dashed lines depict the initial condition. The solid line gives the 1997 tracer inventory in density space, transformed into height z above the target density surface through the mean density profile west of 18° W . The dotted line is the least-squares fit to the solid line from the one-dimensional model.

bottom. Depth-averaged dissipation data from the 1997 survey document a fortnightly modulation in the mixing that lags the intensity of the barotropic tide by a few days (Fig. 5). Microstructure sampling in the vicinity of the main tracer patch occurred chiefly during minima in the fortnightly tidal cycle; this bias resulted in lower dissipation estimates there. Data from both the 1996 and 1997 cruises are biased in these ways, and consequently an earlier basin-average diffusivity estimate⁶ is likely to be an underestimate.

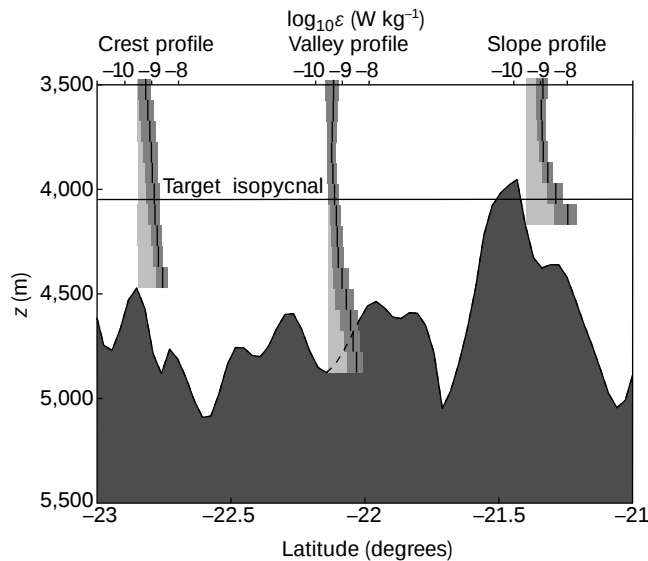


Figure 4 Model profiles of turbulent kinetic energy dissipation rate. Profiles for crests, slopes, and valleys are shown along a section of meridional bathymetry at 18.5° W, the longitude of the tracer injection. The dark grey bands in the profiles indicate 95% confidence intervals about the mean estimates, shown as vertical line segments. The left edge of the light grey band is at $3 \times 10^{-10} \text{ W kg}^{-1}$ for reference. The average dissipation at the injection isopycnal, located at the horizontal line, is smaller over valleys than over crests and slopes.

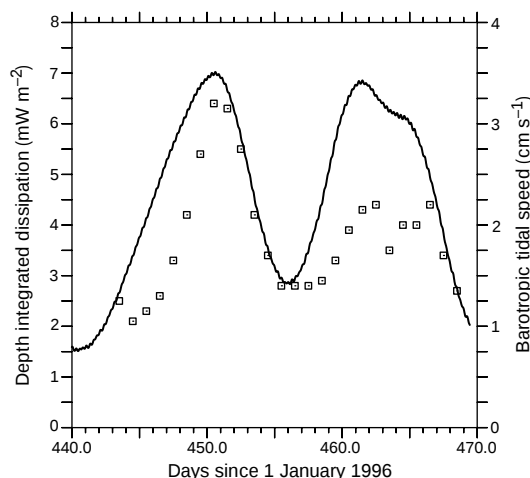


Figure 5 Comparison of the turbulent energy dissipation rate and tidal speeds. The depth-integrated dissipation (squares) and model barotropic tidal speed (solid line: $(u^2 + v^2)^{1/2}$) are shown versus time for the 1997 cruise. A running three-day boxcar filter has been applied to both variables. The tidal velocity estimates¹⁵ were extracted for the times and locations of the stations where the dissipation measurements were made.

The fortnightly cycle in the mixing (Fig. 5) supports the suggestion⁶ that the enhanced mixing in the area results from the breaking of internal waves generated as the barotropic tide flows over the rough ridge bathymetry. Tidal flows over rough bathymetry efficiently generate upward-propagating internal waves with horizontal wavelengths (λ) characteristic of the bottom topography ($\lambda \approx 6\text{--}10 \text{ km}$; ref. 8). Such waves can develop high shear at levels well above the bottom, and this shear can feed turbulent mixing⁹. The fortnightly modulation and small phase lag imply a spatially local balance between internal wave generation by tides and dissipation; the small time lag is consistent with the rate at which bottom-generated internal waves with vertical wavelengths of the order of 100 m can carry energy up into the water column where the waves interact with each other and break.

Both the one-dimensional tracer dispersion model and the dissipation measurements indicate that the diapycnal diffusivity increases sufficiently strongly with depth to imply a divergence of buoyancy flux and therefore that the diapycnal flow must be downward to maintain the density field. For example, the one-dimensional model indicates velocities of -3×10^{-5} to $-5 \times 10^{-5} \text{ cm s}^{-1}$ across isopycnals lying between the ridge tops and the target surface. Close to the bottom, however, there must be a convergence of turbulent buoyancy flux, as, except for a small geothermal flux¹⁰, buoyancy and mass fluxes normal to the bottom must vanish. Hence, there must be a bottom layer in which the diapycnal component of the mean flow is toward decreasing density. This diapycnal flow could be a component of an eastward flow along the bottom. Indeed, a persistent eastward flow below the target density surface is indicated by the tracer plumes, although how much of this flow is along the bottom versus along isopycnals is a subject for further study.

Observations of cold bottom water flowing into the Brazil basin from the south imply upwelling across warmer isotherms and mixing supporting a basin-averaged diffusivity^{4,5} of 2 to 4 $\text{cm}^2 \text{ s}^{-1}$. Based on early results of our experiment extrapolated by bathymetric type to the whole basin, Polzin *et al.*⁶ estimated a heat flux equivalent to an area-averaged diffusivity between 0.5 and 1.5 $\text{cm}^2 \text{ s}^{-1}$ (dictated chiefly by the large dissipation values above the rough Mid-Atlantic Ridge). As the turbulent diffusivities reported here are a factor of two to three larger than those derived from the preliminary observations above the Mid-Atlantic Ridge, a revised estimate now closes the abyssal Brazil basin heat budget within the uncertainties. Moreover, our tracer and turbulence measurements provide guidance on extrapolating to the global ocean overturning problem, in particular by focusing attention on the role of barotropic tides and rough bathymetry in providing energy for mixing through the internal wave field. □

Methods

Tracer data

The average tracer concentration $C(z, t)$ on an isopycnal surface whose mean height is z , was modelled¹¹ by the equation:

$$A \frac{\partial C}{\partial t} + \frac{\partial}{\partial z} [A(w + W)C] = \frac{\partial}{\partial z} \left[AK \left(\frac{\partial C}{\partial z} \right) \right]$$

Here $A(z)$ is the area of the surface at height z within a region encompassing all of the tracer, $w(z)$ is the vertical velocity of the water, and W is an assumed constant vertical velocity of the tracer relative to the water. If the turbulent diffusivities for heat, salt, and tracer are the same, and the equation of state is approximated as linear, then the one-dimensional density equation gives:

$$\left(\frac{\partial z}{\partial t} \right)_\rho + w = \frac{\partial K}{\partial z} + \left(\frac{K}{A} \right) \frac{\partial A}{\partial z} + \left(\frac{K}{N^2} \right) \left(\frac{\partial N^2}{\partial z} \right)$$

Here N is the buoyancy frequency. The first term represents temporal variation of isopycnal depths. Our observations in 1997 and 1998 indicate that this term is small relative to w , justifying its neglect below.

The depth dependence for $K(z)$, chosen with guidance from a model for the internal wave energy source strength⁹, was:

$$K(z) = \frac{K(z_0)}{[1 + \alpha(\sigma(z_0) - \sigma(z))]^2}$$

Here $\sigma(z)$ is the potential density, where $z_0 = 500$ m, which is near the tops of the ridges in the vicinity of the release site and near the bottom of the tracer distribution, and α is a constant. Below z_0 , where there is virtually no tracer at the end of the model period, K was taken to be constant at its value at -500 m, for convenience. No-flux boundary conditions were enforced at $z = -800$ m and $+500$ m. The source or sink of water implicitly balancing $\partial w / \partial z$ carries no tracer, since the concentration is zero at the open side boundaries of the region modelled. The best fit was for $K(z_0) = 8 \text{ cm}^2 \text{ s}^{-1}$ and $\alpha = 12 \text{ m}^3 \text{ kg}^{-1}$, corresponding to $K(0) = 3 \text{ cm}^2 \text{ s}^{-1}$.

Dissipation data

Velocity microstructure data were acquired on the first two cruises with a free-fall instrument¹², deployed in 1997 in conjunction with the tracer sampling (Fig. 1a). From these, turbulent dissipation-rate estimates were derived following Polzin and Montgomery¹³. Adjustment for temporal and spatial biases occurred in three steps. First, temporal bias was addressed by assuming that turbulent dissipation was proportional to the flux of tidally generated internal-wave energy, which, in turn, is proportional to the mean-square tidal current¹⁴. Observed dissipation data (ϵ) were scaled using $\bar{\epsilon}(z) = \{[U^2/U^2(t)]\epsilon(z)$, where $\langle U^2 \rangle$ is the mean-square barotropic current averaged over the period between tracer injection and sampling and $U(t)$ is the tidal current speed for the time of each dissipation estimate. A model¹⁵ was used to estimate the tidal currents. Second, these scaled dissipation data were used to construct model profiles for three bathymetric classes: ridge crests, valleys, and the sloping regions between the two (Fig. 4). Each model profile is a function of local height above bottom. Third, spatial bias was treated by mapping dissipation at the target isopycnal onto a uniform grid with 25-km resolution spanning the area of the tracer patch. The depth of the target isopycnal was mapped using a second-order polynomial in latitude and longitude, fitted to density data in the interval 14–22° W, 19–25° S, and height above bottom was determined with a high-resolution bathymetric data set¹⁶. The appropriate bathymetric class of the dissipation model was used at each grid point and spatial weighting by the observed 1997 tracer concentration field at the target isopycnal was used to calculate the average dissipation. The diffusivity estimate was derived using the expression¹⁷ $K = \Gamma \epsilon N^{-2}$ with the observed mean value of the buoyancy gradient (N^2) and a mixing efficiency (Γ) of 20%. Uncertainty in this estimate derives from stochastic variability in the dissipation and from an uncertainty of 100 m in the bathymetric data. An error analysis was carried out using Monte Carlo methods, and the resulting estimate of $K = 2.3 \pm 1 \text{ cm}^2 \text{ s}^{-1}$ reflects the 68% confidence interval.

Received 25 June; accepted 22 November 1999.

1. Gregg, M. C. Scaling turbulent dissipation in the thermocline. *J. Geophys. Res.* **94**, 9689–9698 (1989).
2. Toole, J. M., Polzin, K. L. & Schmitt, R. W. Estimates of diapycnal mixing in the abyssal ocean. *Science* **264**, 1120–1123 (1994).
3. Polzin, K., Toole, J. & Schmitt, R. Finescale parameterizations of turbulent dissipation. *J. Phys. Oceanogr.* **25**, 306–328 (1995).
4. Hogg, N., Biscaye, P., Gardner, W. & Schmitz, W. J. Jr On transport and modification of Antarctic bottom water in the Vema Channel. *J. Mar. Res. (Suppl.)* **40**, 231–263 (1982).
5. Morris, M., Hogg, N. & Owens, W. B. Diapycnal mixing estimated from advective budgets in the deep Brazil Basin. *Int. WOCE Newsl.* **28**, 23–25 (1997).
6. Polzin, K. L., Toole, J. M., Ledwell, J. R. & Schmitt, R. W. Spatial variability of turbulent mixing in the abyssal ocean. *Science* **276**, 93–96 (1997).
7. Ledwell, J. R., Watson, A. J. & Law, C. S. Mixing of a tracer in the pycnocline. *J. Geophys. Res.* **103**, 21499–21529 (1998).
8. Bell, T. H. Lee waves in stratified flows with simple harmonic time dependence. *J. Fluid Mech.* **67**, 705–722 (1975).
9. Polzin, K. L. Inertial subrange solutions for the energy balance of the finescale internal wavefield. *J. Mar. Res.* (submitted).
10. Langseth, M. G. & Hobart, M. A. Interpretation of heat flow measurements in the Vema Fracture Zone. *Geophys. Res. Lett.* **3**, 241–244 (1976).
11. Ledwell, J. R. & Bratkovich, A. A tracer study of mixing in the Santa Cruz Basin. *J. Geophys. Res.* **100**, 20681–20704 (1995).
12. Schmitt, R. W., Toole, J. M., Koehler, R. L., Mellinger, E. C. & Doherty, K. W. The development of a fine- and microstructure profiler. *J. Atmos. Ocean. Tech.* **5**, 484–500 (1988).
13. Polzin, K. L. & Montgomery, E. T. In *Proc. Microstructure Sensors Workshop* (eds Agrawal, Y., Williams, A. & Goodman, L.) 109–115 (Sequoia Scientific Inc., Mercer Island, Washington 98040, USA, 1997).
14. Bell, T. H., Jr. Topographically generated internal waves in the open ocean. *J. Geophys. Res.* **80**, 320–327 (1975).
15. Egbert, G. D., Bennett, A. F. & Foreman, M. G. G. TOPEX/POSEIDON tides estimated using a global inverse model. *J. Geophys. Res.* **99**, 24821–24852 (1994).
16. Smith, W. H. F. & Sandwell, D. T. Bathymetric prediction from dense altimetry and sparse shipboard bathymetry. *J. Geophys. Res.* **99**, 21803–21824 (1994).
17. Osborn, T. R. Estimates of the local rate of vertical diffusion from dissipation measurements. *J. Phys. Oceanogr.* **10**, 83–89 (1980).

Acknowledgements

Our research is funded by the US National Science Foundation. We thank our technical associates and the officers and crew of the RV *Seward Johnson* for their support.

Correspondence and requests for materials should be addressed to J.R.L. (e-mail: jledwell@whoi.edu).

Osmotic generation of ‘anomalous’ fluid pressures in geological environments

C. E. Neuzil

US Geological Survey, 431 National Center, Reston, Virginia 20192, USA

Osmotic pressures are generated by differences in chemical potential of a solution across a membrane. But whether osmosis can have a significant effect on the pressure of fluids in geological environments has been controversial, because the membrane properties of geological media are poorly understood¹. ‘Anomalous’ pressures—large departures from hydrostatic pressure that are not explicable in terms of topographic or fluid-density effects—are widely found in geological settings, and are commonly considered to result from processes that alter the pore or fluid volume², which in turn implies crustal changes happening at a rate too slow to observe directly. Yet if osmosis can explain some anomalies, there is no need to invoke such dynamic geological processes in those cases. Here I report results of a nine-year *in situ* measurement of fluid pressures and solute concentrations in shale that are consistent with the generation of large (up to 20 MPa) osmotic-pressure anomalies which could persist for tens of millions of years. Osmotic pressures of this magnitude and duration can explain many of the pressure anomalies observed in geological settings. They require, however, small shale porosity and large contrasts in the amount of dissolved solids in the pore waters—criteria that may help to distinguish between osmotic and crustal-dynamic origins of anomalous pressures.

The test was conducted in Cretaceous-age Pierre Shale in central South Dakota, USA. The shale at the site is saturated, has a permeability of 10^{-20} to 10^{-21} m^2 , and is 70–80% clay, of which about 80% is mixed-layer smectite-illite^{3,4}. The shale pore water contains $\sim 3.5 \text{ g l}^{-1}$ NaCl, with only minor amounts of other solutes. Four boreholes were drilled in the shale (Fig. 1), and the test was started by adding waters with different total dissolved solids (TDS) to the boreholes. Fluid squeezed from cores guided the test design, with one borehole receiving a near duplicate of the expelled fluid (designated borehole DUP), two receiving water with ten times the solute concentration (high-TDS boreholes HC1 and HC2), and one receiving deionized water (borehole DI) (see also Supplementary Information, section 1). For 9 years water levels were monitored with an electrical sounding cable, and samples of borehole waters were collected for analysis; the water level and TDS changes are plotted in Fig. 2.

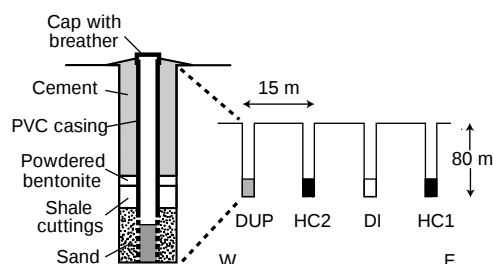


Figure 1 Pierre Shale boreholes used for the osmosis test. The boreholes were capped to prevent rainfall or runoff from entering, but with breather holes to accommodate water level changes. Estimated evaporative losses from barometric pumping through the breather hole during the 9-year test were ~ 1 cm of water. The spacing between boreholes was chosen to ensure that they did not interact during the test.

Figure 2a shows that water levels in HC1 and HC2 rose compared to DUP as osmosis drove water from the shale into the boreholes. The water level in DI initially declined slightly, as osmosis drew water into the shale, and then slowly approached DUP after the TDS difference between DI and DUP became small (Fig. 2b).

Horizontal fluid fluxes in the shale can be described with

$$q = -\frac{k}{\mu} \frac{\partial p}{\partial r} + \sigma \frac{k}{\mu} \frac{\partial \pi}{\partial r} \quad (1)$$

where r (with dimension length, denoted L) is radial distance, q (dimension L^3/L^2T , where T is time) is the porewater flux with respect to the porous matrix, p (dimension M/T^2L , where M is mass) is fluid pressure, k (dimension L^2) is permeability, μ (dimension M/LT) is dynamic viscosity, and σ (dimensionless) is osmotic efficiency⁵. π (dimension M/T^2L) is a measure of the decrease in chemical potential of the water due to the presence of solute, defined as $\pi = -(RT/V_w) \ln(a_w)$, where R (dimension L^2/T^2K) is the gas constant, T (K) is temperature, V_w (dimension L^3/M) is the molar volume of water, and a_w is the water activity⁶. σ , which ranges from 1 to 0, depends strongly on the porewater TDS and the compaction state of the clay in the membrane⁵.

Osmosis causes pore fluid flow (according to the last term in equation (1)) until the resulting pressure build-up creates an equal and opposite hydraulic flow and osmotic equilibrium ($q = 0$) is attained. The fact that the increases in water level in boreholes HC1 and HC2 relative to DUP have slowed significantly, or stopped, (Fig. 2a) is an indication that the test is close to osmotic equilibrium. At equilibrium, equation (1) can be integrated to give

$$\Delta p = \int_{c_{\max}}^{c_{\min}} \sigma(c) \frac{d\pi}{dc} dc \quad (2)$$

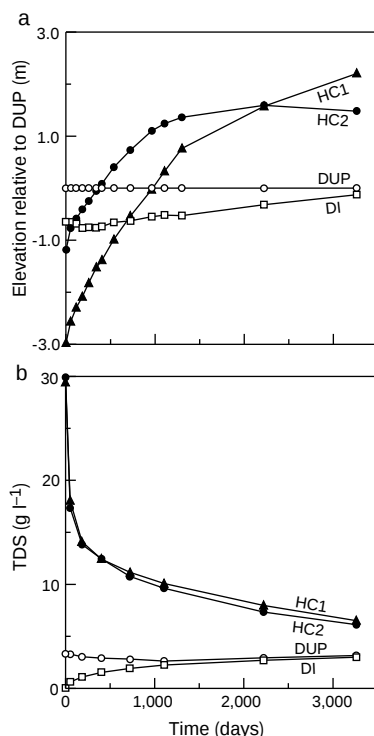


Figure 2 Borehole responses. **a**, Water levels corrected for a regional hydraulic head gradient and plotted as differences from borehole DUP to isolate osmotic responses. **b**, TDS (total dissolved solids) in boreholes. Differences in the borehole diameters caused initial water levels to vary. Osmotic pressure development is typically rapid compared to changes in TDS patterns; in the test, pressure changes were retarded by the fluid storage capacity of the boreholes, with the result that osmotic pressures and TDS changes developed on similar timescales.

where Δp is the pressure build-up and c is the solute concentration (dimension M/L^3). Evaluating Δp requires knowing σ and π as functions of c . π is straightforwardly related to c through water activity a_w , but the relation between σ and c is more complex. Bresler⁵ showed that in clay-based membranes there appears to be a unique relation between σ and $b\sqrt{c}$, where b is a representative half distance between clay platelets. This relation seems to capture accurately the dependence of osmotic efficiency on both solute concentration and compaction state in cakes of smectite, illite and kaolinite. More importantly in the present context, it was also recently found to describe accurately the behaviour of shale cores (B. Cey, personal communication).

Because borehole DUP was presumably unaffected by osmosis, it can be shown that $\Delta p = p - p_{DUP}$, where p_{DUP} is the pressure in DUP. Using concentrations in the borehole and undisturbed shale as limits, I integrated equation (2) numerically and determined b by trial-and-error matching of computed and observed Δp . I then calculated the water content of the shale from its total porosity and equated it to $A_s b$ where A_s (dimension LT^2/M) is the specific surface area of the shale⁷. The resulting A_s value was used to compute b as a function of shale porosity, and Δp could then be computed for any porosity and TDS difference by re-evaluating the integral in equation (2) (see Supplementary Information, section 2). This approach successfully describes osmotic behaviour recently observed in cores of Bearpaw Shale (B. Cey, personal communication), a Pierre Shale equivalent, giving a computed porosity of 0.36 versus a measured value of 0.37 for the cores. In addition, the computed clay particle spacing ($2b$) of ~ 100 Å and specific surface area of ~ 40 m² g⁻¹ are reasonable for the permeability and lithology of the Pierre Shale^{7,8}, suggesting its behaviour *in situ* is unaffected by fractures.

Computed values of Δp are shown in Fig. 3 for the case when the minimum TDS is 10% of the maximum (representing reasonable subsurface conditions), and indicates that the Pierre Shale could generate osmotic pressures reaching ~ 20 MPa (2 km of head) where porosities are small and TDS differences large. More modest pressures of a few megapascals require porosities of about 0.1 or smaller and concentration differences exceeding 20 g l⁻¹, which are common conditions in natural systems.

Figure 3 is probably indicative of the maximum osmotic pressures shales can generate. The high clay content of Pierre Shale, at the 90th percentile for shales⁹ and predominance of

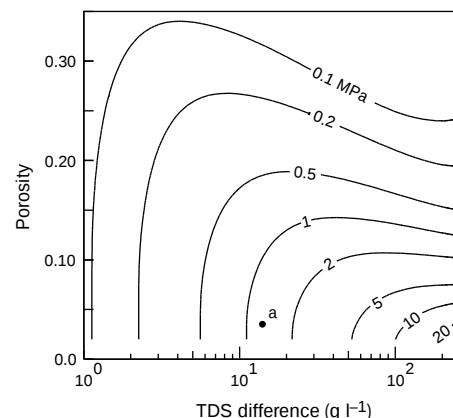


Figure 3 Calculated equilibrium osmotic pressures Δp that could be generated by the Pierre Shale. Minimum TDS was 10% of the maximum, and it was assumed that temperature $T = 350$ K, corresponding to ~ 3 km depth. Point 'a' corresponds to the TDS difference and porosity in a Triassic basin described by Marine and Fritz¹¹ (see text). Increases in osmotic driving forces due to increasing TDS differences can be more than offset by decreases in σ due to higher overall TDS. As a result, osmotic pressure goes through maxima which occur at progressively smaller TDS differences as porosity increases above ~ 0.1 .

smectite-illite should make it an especially efficient geological membrane¹⁰. The data shown in Fig. 3 were calculated assuming that A_s remains constant with compaction, and could overestimate the osmotic pressures that can be generated at very low porosities if loss of clay, or changes in clay mineralogy, accompany compaction. Although there are no test data with which to assess this, one set of field observations provides a point of comparison. In 1981, Marine and Fritz¹¹ published an unusually complete description of a possible osmotic system. In illite-rich shale with a porosity of 0.035, they found a 1.3-MPa pressure anomaly where TDS is elevated $\sim 15 \text{ g l}^{-1}$ above a background of $1\text{--}2 \text{ g l}^{-1}$. When plotted on Fig. 3 (point 'a'), this TDS difference and porosity indicate a pressure similar to the observed value.

Osmotic pressure decays only when TDS differences that maintain it dissipate by diffusion; its longevity is therefore controlled by the membrane's effective ionic diffusion coefficient $\hat{D}_d$ (dimension L^2/T). In the experiment, $\hat{D}_d$ was deduced by analysing TDS changes in the boreholes (Fig. 2b) using:

$$\hat{D}_d \left[\frac{\partial^2 c}{\partial r^2} + \frac{1}{r} \frac{\partial c}{\partial r} \right] = \frac{\partial c}{\partial t} \quad (3)$$

I solved equation (3) for the borehole–shale system, and compared computed and observed TDS changes (see Fig. 4, and section 3 Supplementary Information). TDS in the HC boreholes initially decreased more rapidly than can be accounted for by diffusion. Solubility calculations indicate this resulted from precipitation of mirabilite ($\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$) when the HC solutions cooled after being added to the boreholes. In contrast, the response of borehole DI is reasonably well described by the computed curves, indicating that $\hat{D}_d$ is $(1\text{--}3) \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$. This value will decrease with compaction as σ (and thus ion exclusion) increases, suggesting $10^{-12} \text{ m}^2 \text{ s}^{-1}$ as a representative value at lower porosities. It can be shown that the maximum TDS difference between a shale of thickness L and its surroundings will be halved by diffusion at time $t \approx 0.1L^2/\hat{D}_d$ (ref. 12). Using this as a measure of osmotic pressure longevity with $\hat{D}_d = 10^{-12} \text{ m}^2 \text{ s}^{-1}$, one obtains $t = 1 \text{ Myr}$ if $L = 20 \text{ m}$, and $t \approx 100 \text{ Myr}$ if $L = 200 \text{ m}$, indicating surprisingly long lifetimes for osmotic pressures. Hydrodynamic decay of non-osmotic anomalous pressure (assuming no continuing dynamic process to maintain it) is $\sim 10^3$ times faster, based on the hydraulic diffusivity of Pierre Shale of $\sim 10^{-9} \text{ m}^2 \text{ s}^{-1}$ (ref. 3).

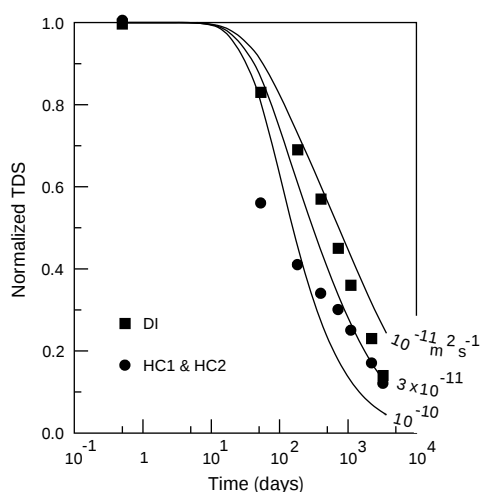


Figure 4 Observed and predicted TDS in boreholes HC1, HC2 and DI. Curves are predicted responses for indicated values of $\hat{D}_d$ in the shale. The vertical scale is $(c(t) - c_s)/(c_0 - c_s)$, where c_s is the background shale TDS, c_0 is the initial borehole TDS, and $c(t)$ is borehole TDS at time t after the start of the test.

The notion of osmotic pressuring in the subsurface has been regarded sceptically¹, despite the possible osmotic system described by Marine and Fritz¹¹ and the fact that many anomalous pressures are associated with areas of high TDS¹³. The results presented here, obtained under *in situ* conditions, are consistent with shales generating osmotic pressures as large as $\sim 20 \text{ MPa}$ and maintaining them for geologically significant periods of time. Osmotic pressures of this magnitude could support a column of water extending 2 km above the land surface, and are sufficient to explain many (though not the largest) observed anomalies. The results also imply that shales may be able to generate high-TDS solutions by retarding solute movement relative to water, a process termed ultrafiltration¹.

The conditions necessary for significant osmotic pressures appear to be common in the subsurface. Most sedimentary basins exhibit large contrasts in solute concentration (often $>200 \text{ g l}^{-1}$; refs 14, 15) and there is now evidence that shales in many of these basins are hydraulically intact over large areas¹⁶. Shales attain the low porosities (~ 0.05 or lower) needed for high membrane efficiency after burial and compaction, processes controlled by basin history. These low shale porosities are typically found below $\sim 1 \text{ km}$ depth in continental basins like the Anadarko and Alberta of North America, where erosion has brought compacted sediments closer to the surface; they are also found below $\sim 5 \text{ km}$ depth in young, active basins like the Gulf Coast or the North Sea which are still receiving sediment¹⁷. Anomalous pressures are frequently encountered in sedimentary basins at similar depths, but have been attributed to continuing and otherwise undetected dynamic processes that include diagenesis, oil generation, and tectonic deformation. It now appears that in many cases the pressure could also be derived osmotically from pre-existing chemical potential differences in the pore water. □

Received 27 July; accepted 17 November 1999.

1. Domenico, P. A. & Schwartz, F. W. *Physical and Chemical Hydrogeology* (John Wiley & Sons, New York, 1990).
2. Ingebritsen, S. E. & Sanford, W. E. *Groundwater in Geologic Processes* (Cambridge Univ. Press, New York, 1998).
3. Neuzil, C. E. Low fluid pressure within the Pierre Shale: A transient response to erosion. *Wat. Resour. Res.* **29**, 2007–2020 (1993).
4. Nichols, T. C. Jr, Collins, D. S. & Davidson, R. R. In situ and laboratory geotechnical tests of the Pierre Shale near Hayes, South Dakota—A characterization of engineering behavior. *Can. Geotech. J.* **23**, 181–194 (1986).
5. Bresler, E. Anion exclusion and coupling effects in nonsteady transport through unsaturated soils: I. Theory. *Soil Sci. Soc. Am. Proc.* **37**, 663–669 (1973).
6. Robinson, R. A. & Stokes, R. H. *Electrolyte Solutions* 2nd edn (Butterworths, London, 1959).
7. Mitchell, J. K. *Fundamentals of Soil Behavior* 2nd edn (John Wiley, New York, 1993).
8. Katsube, T. J. & Williamson, M. A. in *Shales and Mudstones* Vol. II (eds Schieber, J., Zimmerle, W. & Sethi, P. S.) 69–91 (Schweizerbart'sche Verlagsbuch., Stuttgart, 1998).
9. Shaw, D. B. & Weaver, C. E. The mineralogical composition of shales. *J. Sedim. Petrol.* **35**, 213–222 (1965).
10. Fritz, S. J. Ideality of clay membranes in osmotic processes: A review. *Clays Clay Miner.* **34**, 214–223 (1986).
11. Marine, I. W. & Fritz, S. J. Osmotic model to explain anomalous hydraulic heads. *Wat. Resour. Res.* **17**, 73–82 (1981).
12. Crank, J. *The Mathematics of Diffusion* (Oxford Univ. Press, London, 1964).
13. Hanshaw, B. B. & Zen, E.-A. Osmotic equilibrium and overthrust faulting. *Geol. Soc. Am. Bull.* **76**, 1379–1385 (1965).
14. Hanor, J. S. in *Geofluids: Origin, Migration and Evolution of Fluids in Sedimentary Basins* (ed. Parnell, J.) 151–174 (The Geological Society, London, 1994).
15. Bachu, S. Synthesis and model of formation-water flow, Alberta Basin, Canada. *Am. Assoc. Petrol. Geol. Bull.* **79**, 1159–1178 (1995).
16. Neuzil, C. E. How permeable are clays and shales? *Wat. Resour. Res.* **30**, 145–150 (1994).
17. Slater, J. G. & Christie, P. A. F. Continental stretching: An explanation of post-mid-Cretaceous subsidence of the central North Sea Basin. *J. Geophys. Res.* **B 85**, 3711–3739 (1980).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as a paper copy from the London editorial office of Nature.

Acknowledgements

I thank S. Silliman for helping to design the borehole installations; R. Pemberton and L. Matson for drilling and completing the boreholes; B. Baker, C. Solberg, M. Stevens and D. Heyd for measuring water levels and obtaining water samples; and T. Reilly and A. Provost for assisting in the analysis.

Correspondence and requests for materials should be addressed to C.E.N. (e-mail: ceneuzil@usgs.gov).

The most primitive osteichthyan braincase?

Alison M. Basden*, Gavin C. Young†, Michael I. Coates‡ & Alex Ritchie§

* Centre for Ecostratigraphy and Palaeobiology, Department of Earth and Planetary Sciences, Macquarie University, New South Wales 2109, Australia

† Department of Geology, Australian National University, Canberra 0200, Australia

‡ Department of Biological Sciences, University College London, Gower Street, London WC1E 6BT, UK

§ Australian Museum, College St., Sydney South, New South Wales 2000, Australia

Most living vertebrates, from teleosts to tetrapods, are osteichthyans (bony fishes)¹, but the origin of this major group is poorly understood². The actinopterygians (ray-finned bony fishes) are the most successful living vertebrates in terms of diversity. They appear in the fossil record in the Late Silurian but are poorly known before the Late Devonian. Here we report the discovery of the oldest and most primitive actinopterygian-like osteichthyan braincase known, from 400-million-year-old limestone in south-eastern Australia. This specimen displays previously unknown primitive conditions, in particular, an opening for a cartilaginous eyestalk. It provides an important and unique counterpart to the

similarly aged and recently described *Psarolepis* from China and Vietnam^{3,4}. The contrasting features of these specimens, and the unusual anatomy of the new specimen in particular, provide new insights into anatomical conditions close to the evolutionary radiation of all modern osteichthyan groups.

The acid-prepared specimen, AM (Australian Museum) F101607 (Fig. 1) is from the Early Devonian (Emsian, *pireneae-serotinus* conodont zones⁵) Taemas Limestone at Wee Jasper, New South Wales. It consists of an ornamented, anteriorly incomplete skull roof (Fig. 1a), attached to the surface (perichondral) ossification of the cartilaginous braincase. These are preserved in fine detail and show openings for cranial nerves, vessels, sensory canals, and the central cavity for the brain (Fig. 1). The only missing portion is the occipital plate and parachordal region, which presumably was separately ossified, because the posterior face of the preserved portion has a complete perichondral lining, representing continuous ventral and otico-occipital fissures of the braincase^{6,7}. Taemas Limestone fishes are part of the placoderm-dominated Burrinjuck fauna, which also includes lungfishes, other osteichthyan micro-remains and acanthodians⁵. The only other probable actinopterygian described from the fauna is *Ligulalepis toombsi*, known only from scales⁸.

In AMF101607, some bone sutures on the skull roof are obscured by the ornament, but the characteristic osteichthyan pattern of paired frontals and parietal flanked by a series of temporal bones is clear. The only Devonian actinopterygians with comparable braincase preservation are *Mimia toombsi* and *Moythomasia durgaringa*

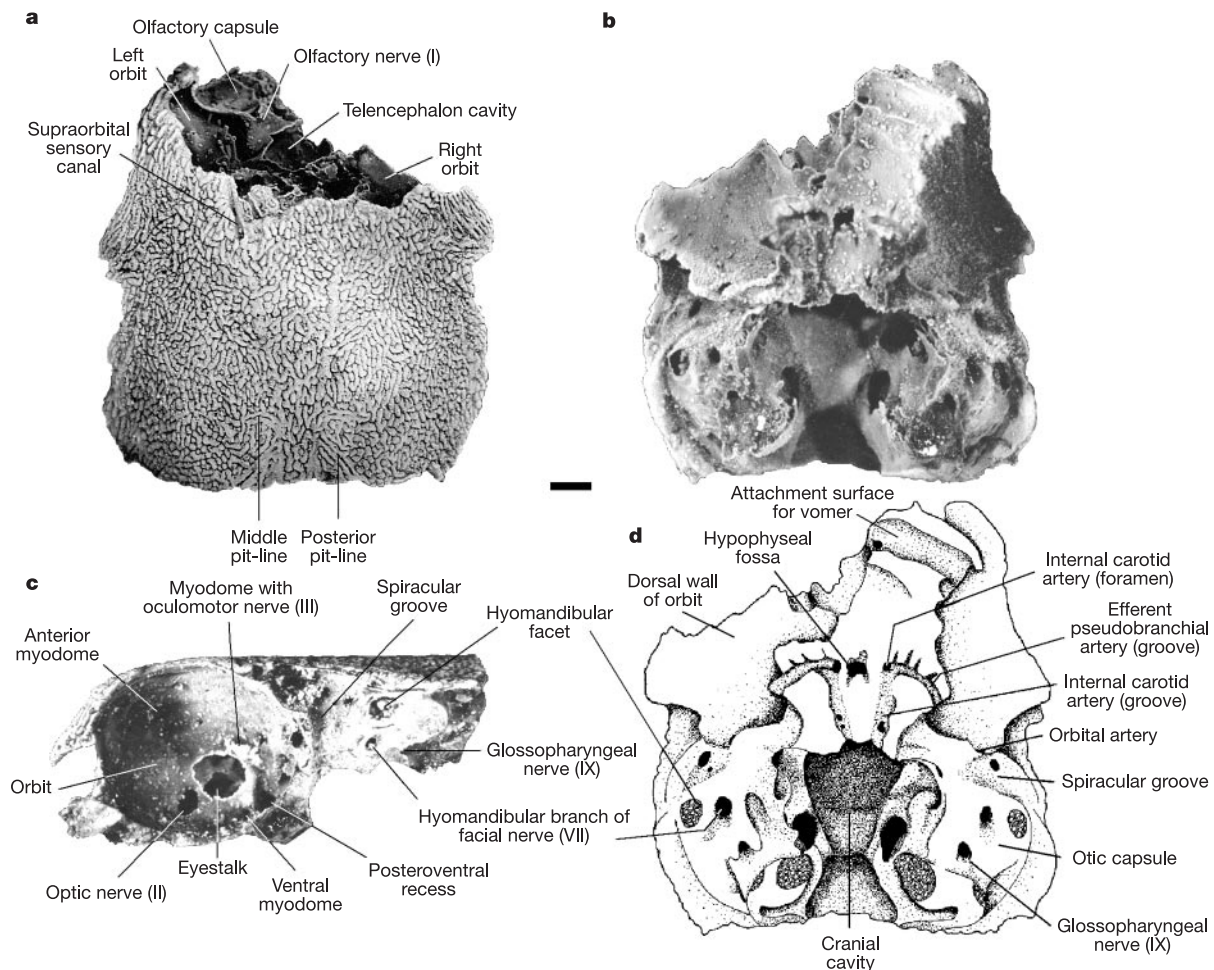


Figure 1 New actinopterygian-like braincase from New South Wales (Australian Museum; AMF101607). **a**, Dorsal view showing skull roof. **b**, Ventral view showing underside of

braincase. **c**, Left lateral view. **d**, Line drawing showing principal features in **b**. Scale bar, 1 mm.

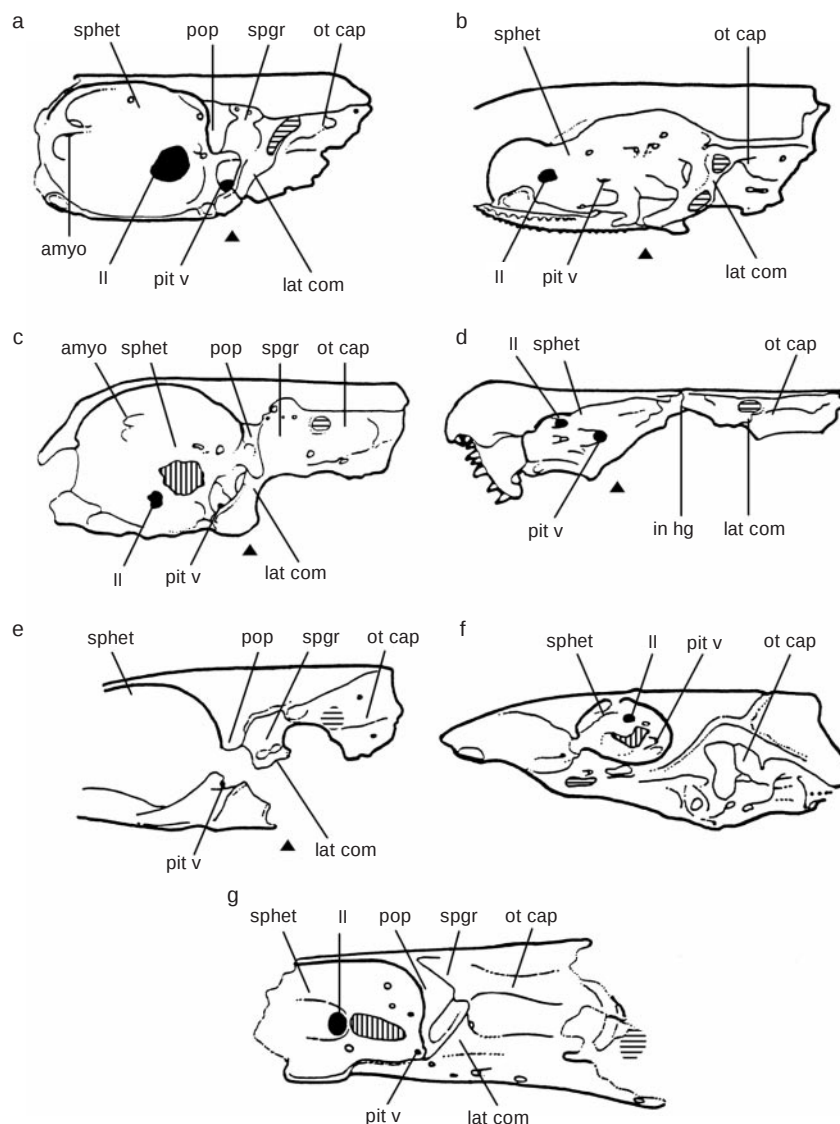


Figure 2 Early braincases of jawed vertebrates (not to scale). All are shown in lateral view with occipital and parachordal regions omitted (except for *Brindabellaspis*) for direct comparison with AMF101607. **a**, *Mimia*⁹, a primitive actinopterygian. **b**, *Youngolepis*²¹, an early sarcopterygian. **c**, AMF101607. **d**, *Psarolepis*³. **e**, *Acanthodes*²⁰, an acanthodian. **f**, *Brindabellaspis*¹¹, a placoderm. **g**, *Cladodus wildungensis*¹⁶, a primitive chondrichth-

yan. Vertical shading shows site of eyestalk; horizontal shading shows articular surface for hyomandibula. Arrow indicates level of ventral cranial fissure. amy, anterior myodome; in hg, intracranial hinge; lat com, lateral commissure; ot cap, otic capsule; pit v, pituitary vein canal; pop, postorbital process; spgr, spiracular groove; sphet, sphenethmoid; II, optic nerve foramen.

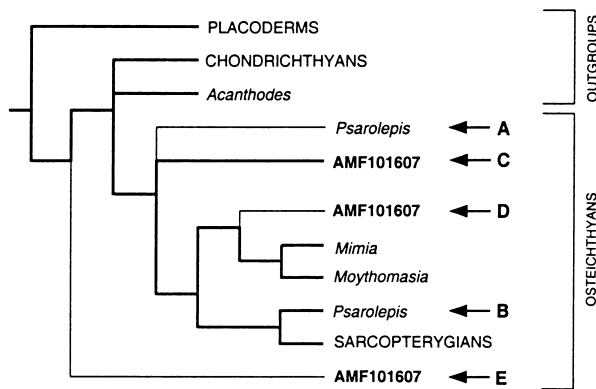
from the Late Devonian of Western Australia⁹. However, many features of AMF101607 are so primitive that comparisons are made more easily with the extinct placoderm fishes. Interpretation of the neuroanatomy in the placoderm braincase is more robust because nerves and vessels can be traced through the thick lateral walls as extensive cranial canals^{10–13}. Although the relationships of placoderms have been controversial^{11,12,14,15}, they probably form an out-group for all the living gnathostomes¹².

The most surprising feature of AMF101607 is the large eyestalk attachment area dominating the medial wall of the orbit (Fig. 1c). Among living vertebrates, eyestalks are restricted to elasmobranchs (sharks and rays), but this form resembles the large eyestalk opening of placoderms^{11–13} (Fig. 2f) or primitive chondrichthyans¹⁶ (Fig. 2g). The eyestalk opening has an everted rim, communicates directly with the interperichondrial space, and therefore must have been cartilage filled in life. In contrast, the margins of the smaller opening directly in front of the eyestalk are rounded and curve inwards to connect with the inner perichondrial lining, thus forming a short canal that transmitted the optic nerve from the cranial

cavity. These conditions also resemble those in placoderms and primitive chondrichthyans (Fig. 2f, g).

Recesses, posterodorsal and ventral to the eyestalk area, indicate probable insertion sites for oculomotor-innervated eye muscles (Fig. 1c), and there is no extensive posterior myodome^{6,17}. Among early actinopterygians, only the Late Devonian *Mimia*⁹ (Fig. 2a) approaches this condition. Instead, the basisphenoid pillar is traversed posteriorly by a slender canal for the pituitary vein. Similar morphologies in placoderms suggest that this condition is primitive for gnathostomes. The ventral recess, which flanks the pillar, communicates through a foramen, probably for the ophthalmic artery, with paired grooves flanking the hypophyseal fossa on the ventral surface of the braincase (Fig. 1b, d). Again, conditions in actinopterygians⁶ and placoderms^{11,12} indicate that this arrangement is primitive. However, like chondrichthyans and (other) osteichthyans, but unlike placoderms, the superior oblique eye muscle probably inserted in the anterodorsal wall of the orbit (anterior myodome, Fig. 1c).

A prominent postorbital process (pop, Fig. 2c) divides orbital



	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>Norselaspis</i>	0	0	0	?	0	0	0	0	0	0	0	0	0
Placoderms	0	0	0	1	0	0	1	0	0	0	0	0	0
Chondrichthyans	0	2	0	1	0	0	1	0	0	0	0	1	1
<i>Acanthodes</i>	0	1	0	?	0	0	1	0	0	0	0	?	1
<i>Mimia</i>	1	2	0	?	0	0	1	1	0	1	2	1	1
<i>Moythomasia</i>	1	2	0	?	1	1	1	1	1	1	2	1	1
AMF101607	1	1	0	1	0	0	1	0	0	0	1	1	1
<i>Psarolepis</i>	2	?	1	?	0	0	0	1	0	0	1	1	?
Sarcopterygians	2	2	1	0	0	0	0	1	0	0	1	1	2

Figure 3 Tree topology of Zhu *et al.*⁴ showing alternative positions for two contrasting primitive osteichthyan braincases, AMF101607 and *Psarolepis*^{3,4}. An identical tree topology to that indicated by bold lines can be obtained from PAUP²² analysis of the data matrix of cranial characters. **1**, rectilinear skull roof pattern: absent, 0; undivided, 1; divided, 2. **2**, otic cranial fissures: absent, 0; present and confluent, 1; present and separate, 2. **3**, intracranial hinge: absent, 0; present, 1. **4**, eyestalk: absent, 0; present, 1. **5**, postorbital process dividing ethmosphenoid from otic braincase regions: absent, 0; present, 1. **6**, hyomandibular articulates on rear of lateral commissure: absent, 0; present, 1. **7**, spiracular groove open, 0; bridged/canalized, 1. **8**, pituitary canal: narrow, 0; broad, 1. **9**, posterior myodome roof pierced by canal communicating with postorbital intramural space: absent, 0; present, 1. **10**, narrow interorbital space: absent, 0; present, 1. **11**, pit-lines on actinopterygian parietal/sarcopterygian postparietal: absent, 0; two, 1; three, 2. **12**, one or both oblique eye muscle insertions anterior to optic nerve foramen: absent, 0; present, 1. **13**, hyomandibular articular facets: absent, 0; single, 1; double, 2.

(sphenethmoid) from otic braincase regions, and a broad spiracular groove traverses the posterolateral surface of this process. Below this process, the groove passes along an ossified lateral commissure¹⁸, bridging the jugular canal and joining the basisphenoid to the anterolateral face of the otic capsule (Fig. 1c, d). This resembles actinopterygian, acanthodian and primitive chondrichthyan conditions¹⁹ (Fig. 2a, c, e, g), rather than the more posteriorly sited commissure of sarcopterygians (Fig. 2b, d).

In primitive osteichthyans, the hyomandibula (homologue of the tetrapod stapes) articulates with posterolaterally directed facets on the rear of the postorbital process and/or lateral commissure (Fig. 2a, b, d). In actinopterygians, this articular surface is usually dorsoventrally elongate. AMF101607 is thus the only primitive osteichthyan with a small, subcircular hyomandibular articular facet, with a roughened surface, located half-way along the anteroposterior axis of the otic capsule wall (Fig. 1b–d). This places it level with the inferred articular site in *Acanthodes*²⁰ (Fig. 2e) and about half-way between conventional primitive osteichthyan and chondrichthyan patterns (Fig. 2g), in which the hyomandibula attaches to the posterolateral angle of the otic capsule¹⁵. There is no evidence that the postorbital process articulated directly with any part of the mandibular arch.

AMF101607 resembles *Acanthodes* and primitive actinopterygians in lateral view (Fig. 2a, e), but dorsal and ventral aspects (Fig. 1a, b, d) show several sarcopterygian-like features. The paths of

the supraorbital sensory canals converge posteriorly on a pair of curved pit lines (Fig. 1a), which resemble those of *Psarolepis*³ rather than the clusters of three straight lines present in primitive actinopterygians such as *Mimia* and *Moythomasia*⁹. The short, broad cavity of the telencephalon, exposed beneath the broken anterior of the skull roof (Fig. 1a), lies between widely separated orbital walls, whereas in other actinopterygians these walls are in near or actual contact^{6,9}. In ventral view (Fig. 1b, d), the lateral span of the basisphenoid is quite unlike any known actinopterygian, and can only be compared to sarcopterygian conditions, again exemplified by *Youngolepis*²¹ and *Psarolepis*³.

The detail preserved in this specimen provides a new character mosaic for cranial conditions in primitive osteichthyans. The phylogenetic impact of these data has yet to be fully explored, but certain consequences are readily apparent. For example, the presence of an eyestalk supports the contention that this feature is primitive for gnathostomes⁹, and so can no longer represent a synapomorphy uniting placoderms and chondrichthyans, as previously suggested^{12,13}. Acanthodians are now the only major group in which an eyestalk has yet to be found.

If these and other features of AMF101607 are summarized as character states and optimized on the tree topology presented by Zhu *et al.*⁴, they can be used to reassess the alternative positions considered for *Psarolepis* (Fig. 2d; for character list and matrix, see Fig. 3), which has been interpreted as either a basal osteichthyan or a basal sarcopterygian (positions A and B, Fig. 3). Placing AMF101607 as the actinopterygian sister group (position D, Fig. 3) requires a tree length of 22 steps (braincase-related characters) to accommodate *Psarolepis* in position A (Fig. 3), instead of 20 steps for position B. *Psarolepis* is thus more likely to be a primitive sarcopterygian, as originally reported³, on the basis of the following synapomorphies: division of the rectilinear roof pattern (character 1, state 2); presence of an intracranial hinge (character 3, state 1); and absence of a clearly defined postorbital process (character 7, state 0). These characters do not always occur together (see ref. 21 or Fig. 2b), but are probably inter-related because of structural and functional specializations in primitive sarcopterygian crania. Taking a broader view, and given the range of primitive characters present, it is worth considering the possibility that AMF101607 branches from the tree between placoderms and chondrichthyans (position E, Fig. 3). This solution requires only 20 steps. However, the simplest interpretation of these data is to retain *Psarolepis* in position B (Fig. 3) as a basal sarcopterygian, but move AMF101607 to position C (Fig. 3) as a basal osteichthyan. The resultant tree length is only 19 steps, and a branching sequence identical to that obtained by analysing the character matrix using PAUP²². We recognize that any conclusion based on this result must be treated as tentative, but it points to the intriguing possibility that AMF101607 is not just a primitive actinopterygian, but a stem-group osteichthyan. If this is correct, then AMF101607 reveals primitive cranial conditions for not only the lineage ancestral to ray-finned fish, but also the lineage leading to lobe-finned fishes and ourselves. □

Received 8 June; accepted 22 October 1999.

- Janvier, P. *Early Vertebrates*. Oxford Monographs on Geology and Geophysics Vol. 33 (Oxford Univ. Press, Oxford, 1996).
- Ahlberg, P. E. Something fishy in the family tree. *Nature* **397**, 564–565 (1999).
- Yu, X. A new porolepiform-like fish, *Psarolepis romeri*, gen et sp. nov. (Sarcopterygii, Osteichthyes) from the Lower Devonian of Yunnan, China. *J. Vert. Paleontol.* **18**, 261–274 (1998).
- Zhu, M. Yu, X. & Janvier, P. A primitive fossil fish sheds light on the origin of bony fishes. *Nature* **397**, 607–610 (1999).
- Basden, A. M., Hocking, M., Parkes, R., Burrow, C. J. & Young, G. C. Siluro-Devonian microvertebrates from southeastern Australia. *Cour. Forsch. Senck.* (in the press).
- Patterson, C. The braincase of pholidophorid and leptolepid fishes, with a review of the actinopterygian braincase. *Phil. Trans. R. Soc. Lond. B* **269**, 275–579 (1975).
- Gardiner, B. G. & Bartram, A. W. H. in *Problems in Vertebrate Evolution* (eds Andrews, S. M., Miles, R. S. & Walker, A. D.) 227–245 (Academic, London, 1977).
- Schultze, H.-P. Palaeoniscoidea-Schuppen aus dem Unterdevon Australiens und Kanadas und aus dem Mitteldevon Spitsbergens. *Bull. Brit. Mus. Nat. Hist. (Geol)* **16**, 343–368 (1968).

9. Gardiner, B. G. The relationships of the palaeoniscid fishes, a review based on new specimens of *Mimia* and *Moythomasia* from the Upper Devonian of Western Australia. *Bull. Brit. Mus. Nat. Hist. (Geol.)* **37**, 173–428 (1984).
10. Young, G. C. New information on the structure and relationships of *Buchanosteus* (Placodermi: Euarthrodira) from the Early Devonian of New South Wales. *Zool. Jour. Linn. Soc.* **66**, 307–352 (1979).
11. Young, G. C. A new early Devonian placoderm from New South Wales, Australia, with a discussion of placoderm phylogeny. *Palaeontographica* **167A**, 10–76 (1980).
12. Young, G. C. The relationships of placoderm fishes. *Zool. Jour. Linn. Soc.* **88**, 1–57 (1986).
13. Goujet, D. Placoderm interrelationships: a new interpretation, with a short review of placoderm classification. *Proc. Linn. Soc. NSW* **107**, 211–243 (1984).
14. Goujet, D. Les affinités des Placodermes, une revue des hypothèses actuelles. *Géobios, Mém. Spéc.* **6**, 27–38 (1982).
15. Gardiner, B. G. The relationships of placoderms. *J. Vert. Paleont.* **4**, 379–395 (1984).
16. Gross, W. Das Kopfskelett von *Cladodus wildungensis* Jaekel. *Senckenbergiana* **19**, 80–107 (1937).
17. Schaeffer, B. & Dalquest, W. W. A palaeonisciform braincase from the Permian of Texas, with comments on cranial fissures and the posterior myodome. *Am. Mus. Novit.* **2658**, 1–15 (1937).
18. de Beer, G. R. *The Development of the Vertebrate Skull* (Oxford Univ. Press, Oxford, 1937).
19. Coates, M. I. & Sequeira, S. E. K. The braincase of a primitive shark. *Trans. R. Soc. Edinb., Earth Sci.* **89**, 63–85 (1998).
20. Miles, R. S. in *Interrelationships of Fishes* (eds Greenwood, P. H., Miles, R. S. & Patterson, C.) *Zool. J. Linn. Soc.* **53** (suppl. 1), 63–103 (1973).
21. Chang, M. M. The braincase of *Youngolepis*, a Lower Devonian crossopterygian from Yunnan, south-western China. Thesis, Univ. Stockholm (1982).
22. Swofford, D. L. PAUP: phylogenetic analysis using parsimony, version 3.1.1. (Illinois Natural History Survey, Champaign, Illinois, 1993).

Acknowledgements

We thank D. Goujet and C. Poplin for helpful discussion. A.B. thanks the Australian Museum for use of facilities, and acknowledges support from an Australian Postgraduate Award and Macquarie University Postgraduate Fund. M.I.C. acknowledges support from the Biotechnology and Biological Sciences Research Council UK.

Correspondence and requests for materials should be addressed to A.B. (e-mail: abasden@laurel.ocs.mq.edu.au).

Molecular evidence regarding the origin of echolocation and flight in bats

Emma C. Teeling^{*†}, Mark Scally^{*†}, Diana J. Kao[‡], Michael L. Romagnoli[‡], Mark S. Springer[‡] & Michael J. Stanhope^{*†}

^{*}Queen's University of Belfast, Biology and Biochemistry, 97 Lisburn Road, Belfast BT9 7BL, UK

[‡]Department of Biology, University of California, Riverside, California 92521, USA

Bats (order Chiroptera) are one of the few orders of mammals that echolocate and the only group with the capacity for powered flight. The order is subdivided into Microchiroptera and Megachiroptera, with an array of characteristics defining each group¹, including complex laryngeal echolocation systems in microbats and enhanced visual acuity in megabats. The respective monophylies of the two suborders have been tacitly assumed, although microbat monophyly is uncorroborated by molecular data. Here we present a phylogenetic analysis of bat relationships using DNA sequence data from four nuclear genes and three mitochondrial genes (total of 8,230 base pairs), indicating that microbat families in the superfamily Rhinolophoidea are more closely related to megabats than they are to other microbats. This implies that echolocation systems either evolved independently in rhinolophoids and other microbats or were lost in the evolution of megabats. Our data also reject flying lemur (order Dermoptera) as the bat sister group, indicating that presumed shared derived

characters for flying lemurs and bats² are convergent features that evolved in association with gliding and flight, respectively.

Bat species were chosen to include at least one representative from each microbat superfamily³ as well as representative megabats. Because the choice of outgroup can have a major influence on reconstructed ingroup topologies, and because the sister group to bats is controversial^{4,5}, we included four diverse outgroup taxa. We obtained new sequences and/or extracted sequences from GenBank for four nuclear genes and three mitochondrial genes. The nuclear genes were exon 11 of BRCA1 (breast cancer susceptibility gene; 2.8 kilobases (kb)), exon 28 of vWF (von Willebrand factor; 1.2 kb), RAG1 (recombination activating gene 1; 1.1 kb), and RAG2 (recombination activating gene 2; 0.8 kb). The mitochondrial genes were the complete 12S and 16S ribosomal RNA genes and the intervening valine transfer RNA gene. Independent mitochondrial and nuclear loci were chosen to index different types of DNA sequence information.

None of our phylogenetic analyses supported the monophyly of microbats. We found that *Megaderma* and *Hipposideros*, two microbat species from different families within the superfamily Rhinolophoidea, grouped as the sister clade to the megabats (Fig. 1). Individual loci did not provide decisive bootstrap support for this association, but all loci provided higher support for Rhinolophoidea plus megabats than for microbat monophyly. For example, with neighbour-joining and logdet distances (all sites included), the bootstrap support deriving from independent genes for Rhinolophoidea plus megabats and microbat monophyly, respectively, was as follows: 12S–16S: 55 and 10; RAG1: 54 and 0; RAG2: 37 and 15; vWF: 25 and 11; BRCA1: 24 and 1. Because none of the strongly supported clades based on individual data sets were mutually incompatible, we concatenated the sequences into combined nuclear and mitochondrial-plus-nuclear data sets. The main principle behind combining data is that it allows for amplification of phylogenetic signal, and increased resolving power, in cases where signal is masked by homoplasy among the individual data sets. Analyses based on concatenated nuclear and mitochondrial-plus-nuclear data sets provided much higher bootstrap support for Rhinolophoidea plus megabats (Fig. 1; Table 1) and statistical tests with both parsimony and maximum likelihood rejected microbat monophyly in favour of the rhinolophoid–megabat alliance (Table 2).

Furthermore, using the variable-length bootstrap^{6,7}, support for Rhinolophoidea plus megabats increased as a function of the number of resampled base pairs for each locus (Fig. 2a), except for BRCA1. Removing third positions of BRCA1 resulted in the same pattern of increasing support for the Rhinolophoidea–megabat hypothesis (Fig. 2a). Third positions of BRCA1 exhibit two characteristics, base compositional heterogeneity and lineage specific rate variation, that may confound phylogenetic estimations. With or without third positions, BRCA1 did not support microbat monophyly. Bootstrap support with the concatenated data sets increased as a function of the number of re-sampled base pairs, much as for the individual genes (Fig. 2b), but was always higher with similar numbers of nucleotides sampled from the concatenated alignment compared to the individual genes (compare Fig. 2a and b, respectively). This finding parallels results⁸ wherein individual mitochondrial genes did not perform as well as random samples of equivalent size that were drawn from the pooled data set. Also, much of the increased resolving power of the pooled data relative to the individual data sets is because of the increased number of base pairs in the pooled data. For many problems in the molecular phylogenetic analysis of mammalian evolutionary history, it is likely that adequate resolving power will only be achieved through data set concatenation; the present finding is simply one example.

Phylogenetic analyses can yield results that are artefacts of the choice of outgroup, long-branch attraction, heterogeneous base composition and site-specific rate variation⁹. We therefore investi-

[†]Present addresses: Department of Biology, University of California, Riverside, California 92521, USA (E.C.T., M.S.); Bioinformatics, SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, UP1345, Collegeville, Pennsylvania 19426-0989, USA (M.J.S.).

gated whether any of these factors contributed to the phylogenetic association of rhinolophoids and megabats. Outgroup choice sensitivity analyses consistently resulted in strong bootstrap support for rhinolophoids plus megabats, irrespective of which taxon was chosen as the outgroup (57%–100% bootstrap support, with 10 of the 16 different analyses performed in excess of 85%). Phylogenetic analyses were also performed with a 10-taxon data set that excluded *Tonatia* and *Mus* because our examination of these data indicated potential problems with long edges and/or base composition involving these two taxa (Table 1). As before, the association of rhinolophoids and megabats was strongly supported (Table 1). Statistical tests with the 10-taxon data set rejected microbat monophyly (Table 2). Finally, maximum likelihood bootstrap analyses that accounted for site-specific rate variation with a gamma distribution of rates provided 100% bootstrap support for the rhino-

lophoid–megabat association with both the 12-taxon and 10-taxon concatenations of mitochondrial and nuclear genes (Table 1).

Our phylogenetic result is also supported by other types of data. Hutcheon *et al.*¹⁰ suggested that rhinolophoids and megabats are sister taxa based on single-copy DNA hybridization results. Rhinolophoids and megabats also share a high AT bias and an alternative interpretation of the hybridization data was that rhinolophoids and megabats clustered together as an artefact of base composition¹⁰. For the genes that we sequenced, we did not find evidence for high AT content in megabats and rhinolophoids relative to other chiropterans. For example, among informative sites in our alignment, the percentage of A + T was higher in vespertilionoids (53.5%–56.4%), megabats (52.9%–55.3%) and the emballonurid (55.3%) than in rhinolophoids (48.9%–52.3%) or the phyllostomid (44.2%). These results are inconsistent with an association of

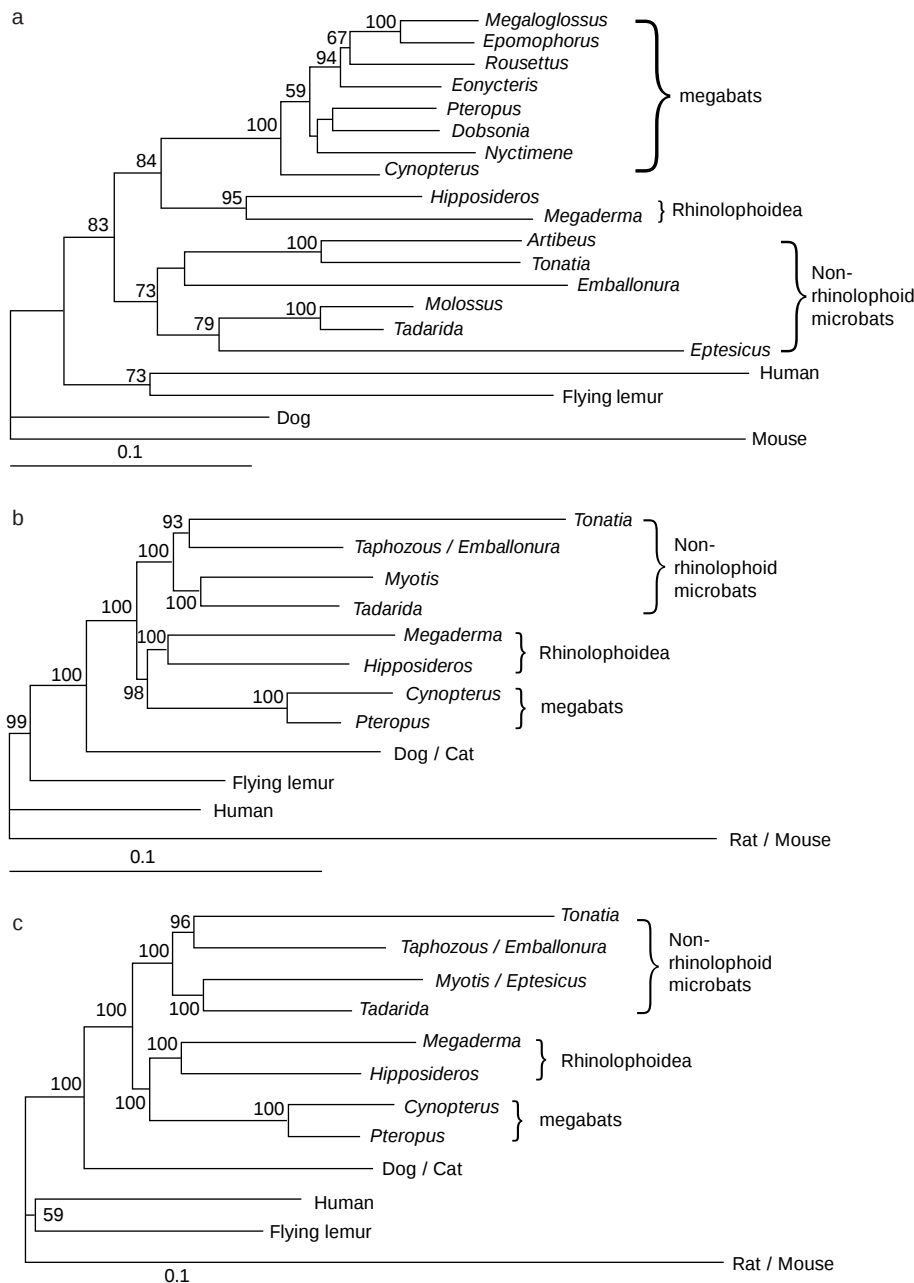


Figure 1 Maximum likelihood trees with branch lengths drawn proportional to amount of sequence change and with maximum likelihood bootstrap figures indicated. **a**, 12S–16S; **b**, combined nuclear genes; **c**, combined nuclear and mitochondrial sequences. **b,c**, A

solidus separating two taxonomic names indicates a hybrid sequence involving those two taxa.

Table 1 Bootstrap support for Rhinolophoidea-megabat and microbat monophyly using indicated phylogenetic methods and data positions

	Parsimony		Transversion parsimony		Minimum evolution-Logdet		Minimum evolution-maximum likelihood		Maximum likelihood with gamma		Maximum likelihood	
	RM	MM	RM	MM	RM	MM	RM	MM	RM	MM	RM	MM
12 taxa, nuclear-plus-mitochondrial	98	2	90	1	99	1	100	0	100	0	100	0
20 taxa, mitochondrial	73	11	44	3	58	9	68	5	89	6	77	9
12 taxa, nuclear	90	3	89	9	97	1	98	1	98	1	96	1
10 taxa, mitochondrial	66	8	42	0	50	7	53	4	94	0	79	5
10 taxa, nuclear	62	3	91	8	92	5	89	9	99	0	88	12
10 taxa, nuclear-plus-mitochondrial	91	1	70	1	98	2	97	2	100	0	100	0

RM, Rhinolophoidea-megabats; MM, microbat monophyly. The rationale behind conducting a 10-taxon analysis involved the following: Relative Apparent Synapomorphy Analysis (RASA)²⁶ indicated lack of significant phylogenetic signal for the 12-taxon, mitochondrial-plus-nuclear data set; such a result can be an indicator of long-branch attraction²⁶. Removal of *Tonatia* and *Mus*, the ingroup and outgroup taxa with the longest edges, respectively, resulted in significant phylogenetic signal at $P = 0.001$ (tRASA = 4.699). Also, a chi-square test of homogeneity in base composition (variable sites only) was significant when all 12 taxa were included ($P = 0.00000000$), but not significant when *Tonatia* was eliminated ($P = 0.13$). Logdet values shown above are based on analyses with both variable and constant sites. With the removal of constant sites in the logdet analyses, bootstrap support for the Rhinolophoidea-megabat association with the combined nuclear-plus-mitochondrial data sets was 96% for both the 12-taxon and 10-taxon analyses.

rhinolophoids and megabats that is merely an artefact of high A + T content. Also, results obtained with the logdet method, which addresses potential concerns with heterogeneity of base composition⁹, agree with results for other methods in supporting a rhinolophoid-megabat alliance (Table 1). With logdet, bootstrap support for Rhinolophoidea plus megabats with the combined data sets always exceeded 90%, even with the removal of constant sites (Table 1).

Beyond molecular evidence favouring a rhinolophoid-megabat alliance, there are other considerations that support this hypothesis. First, a sister-group relationship between rhinolophoids and megabats, rather than between megabats and all other chiropterans, reduces the duration of the implied ghost lineage for megabats from the early Eocene (the holotype of the earliest bat, *Icaronycteris*, is dated at 53 million years ago¹¹) to the late Eocene (there are fossil rhinolophoids at 37–34 million years ago^{4,12}). Another non-molecular consideration that supports a rhinolophoid-megabat association is that both groups have distributions that are restricted to the Old World.

Our results have implications for the evolution of both flight and echolocation in bats. First, bats and flying lemurs did not share a flying common ancestor as has been suggested in various morphological analyses conducted over the course of the last century^{13,14}. None of our analyses support a sister-group relationship between bats and flying lemur, and indeed this hypothesis is rejected in statistical tests involving the concatenated genes (Table 2). Instead, we find, in agreement with other molecular studies⁵, that bats are not part of a monophyletic group that includes other archontans (such as primates and flying lemurs). Although it is likely that the ancestor of bats progressed through an intermediate flying stage¹², we suggest that this must have occurred independently of the evolution of flight in flying lemurs. Numerous presumed synapomorphies for bats and flying lemurs², including markedly elongated forelimbs, presence of a humeropatagialis muscle and fusion of several carpal elements into the scaphocentralunate, are convergent features that evolved in association with gliding or flight. Such an interpretation would concur with another morphological analysis¹⁵ which argued against Volitantia and favoured a Primates/Dermoptera clade.

Second, our phylogeny indicates that echolocation either evolved in the ancestor of crown group bats and was then lost in Old World fruitbats or evolved independently in rhinolophoids versus other microchiropterans. Regarding the first scenario, some workers have argued for the implausibility of loss of laryngeal echolocation because echolocation has obvious benefits, with only minimal cost, when coupled with flight^{3,16}. We suggest that changes in roosting locations and foraging habits may have obviated the need for echolocation. Also, laryngeal echolocation systems may limit body size because echolocation pulse frequency is coupled to

wingbeat rate and ventilation rate. If wingbeat rate, ventilation rate, and pulse frequency became too low, selection for increased body size in an ancestral megabat may have compromised the selective advantage of echolocation. Laryngeal echolocation may have been selected against in favour of enhanced visual acuity and larger body size.

As for the second scenario, several lines of evidence suggest that echolocation may have evolved independently among different

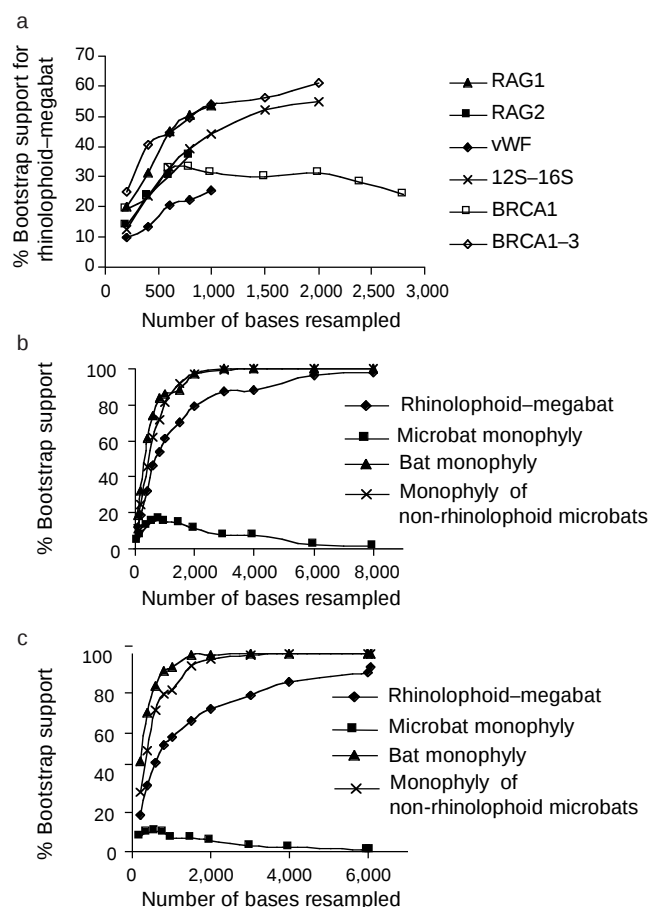


Figure 2 Bootstrap support using neighbour-joining with logdet distances, for various phylogenetic associations. Percentage values are given as a function of the number of resampled nucleotides, and there are various data partitions. **a**, Rhinolophoidea-megabat support for each of the individual loci. **b**, Support for various possible associations involving the bats, using the concatenated nuclear gene data set. **c**, Support for bat phylogenetic associations involving the nuclear-plus-mitochondrial data set.

Table 2 Results of statistical tests on microbat monophyly and the Volitantia concept

	Additional steps to maximum parsimony tree*	Decrease in likelihood to maximum likelihood tree†	Kishino Hasegawa test parsimony	Templeton test parsimony	Kishino Hasegawa test maximum likelihood
Microbat monophyly					
12-taxon analyses	+15	24.511; s.d. = 9.445	0.0394	0.0394	0.0095
10-taxon analyses	+14	36.528; s.d. = 13.592	0.0433	0.0433	0.0072
Volitantia					
12-taxon analyses	+65	106.063; s.d. = 19.355	0.0001	0.0001	<0.0001
10-taxon analyses	+82	154.461; s.d. = 23.638	<0.0001	<0.0001	<0.0001

* Score is 7,128 for 12 taxa and 5,189 for 10 taxa.

† -lnL = 44,397.611 for 12 taxa and 36,346.401 for 10 taxa.

s.d. standard deviation.

microbats. Echolocation is known to occur among other mammalian groups including cetaceans and insectivores, as well as the megabat *Rousettus*^{12,17–19}. The cost of laryngeal echolocation is relatively low when coupled with flight²⁰. Microbat echolocation systems are highly variable, but nonetheless taxon-specific, illustrating the evolutionary flexibility of bat echolocation. Some bats emit echolocation pulses orally whereas in other taxa they are emitted through the nasal cavity. In addition, echolocation pulses range from frequency-modulated to constant-frequency¹². There are numerous putative morphological synapomorphies for microbats. However, many such presumed synapomorphies of microchiropterans, including features related to flight and ventilation, are either directly related to echolocation or are functionally correlated with the echolocation system. Thus, they do not represent independent evolutionary events in support of microbat monophyly. For example, the enlarged orbicular process on the malleus, common to all microbats, may help to avoid self-deafening and/or improve the ability of the middle-ear ossicles to transmit high-frequency sounds with minimal delay time⁴. Features of the postcranial skeleton in microbats include anterior laminae and enlarged posterior laminae on the ribs. Rib laminae stiffen the ribcage and provide a greater area for muscle origins, for example, muscles of the serratus anterior complex. In microbats, there is a close association between wingbeat and sound emission²¹; Simmons and Geisler⁴ hypothesized that the anterior and posterior laminae on the ribs, along with the expanded anterior serratus, increase the efficiency of the flight mechanism, ventilation and the production of echolocation calls. Rather than being synapomorphic for all microchiropterans, features such as expanded rib laminae may have evolved independently in rhinolophoids and other microbats in conjunction with the evolution of laryngeal echolocation. □

Methods

Data collection and taxa representation

Taxon sampling was designed to include at least one representative (in some cases two) of the four proposed microbat superfamilies³, as well as megabat representatives, combined with diverse outgroups, for all loci. The microbat taxa and their loci were: (1) superfamily Emballonuroidea, family Emballonuridae: *Taphozous* sp. (BRCA1, RAG1, RAG2); *Emballonura atrata* (12S–16S, vWF). (2) Rhinolophoidea, Megadermatidae: *Megaderma lyra* (all loci); Hipposideridae: *Hipposideros commersoni* (all nuclear loci), *Hipposideros galeritis* (12S–16S). (3) Phyllostomoidea, Phyllostomidae: *Tonatia bidens* (all loci). (4) Vespertilionoidea, Vespertilionidae: *Myotis daubentonii* (BRCA1, RAG1, RAG2), *Myotis velifer* (vWF), *Eptesicus fuscus* / *capensis* (12S–16S); Molossidae: *Tadarida brasiliensis* (all loci), *Molossus sinaloae* (12S–16S). Megabats included the following taxa: Pteropodidae: *Cynopterus sphinx* (all nuclear loci), *Cynopterus brachyotis* (12S–16S), *Pteropus rayneri* (BRCA1, RAG1, RAG2), *Pteropus hypomelanus* (12S–16S, vWF) and 12S–16S from *Nyctimene robinsoni*, *Dobsonia moluccensis*, *Megaloglossus woermanni*, *Epomophorus wahlbergi*, *Rousettus amplexicaudatus* and *Eonycteris spelaea*. Outgroup taxa included the following: flying lemur (*Cynocephalus variegatus*, Dermoptera; all loci), human (*Homo sapiens*, Primates; all loci), mouse (*Mus musculus*, Rodentia; all loci except vWF), rat (*Rattus norvegicus*, Rodentia; vWF), dog (*Canis familiaris*, Carnivora; all loci except RAG1 and RAG2) and cat (*Felis catus*, Carnivora; RAG1 and RAG2). Accession numbers for the 10 mitochondrial and 32 nuclear sequences new to this study are: AF203737–AF203778. To obtain the complete alignments these sequences were combined with those already available—mitochondrial: AF203726, AF17929, AF179288, AF061340, U61082, U93053–5, U93059, U93061, U93064–5, U93068, U93070, U97073, AF069536–8, AF044606–7,

AF044610–11, AF044620, AF044626, J01420, U96639, J01415; nuclear: M29474–5, M94633, M64796, U31616, U31622, AF061061, U31605–6, L76227, M25851, AJ224673, U68174, AF019081, U50709, L78833.

Polymerase chain reaction (PCR) and sequencing for 12S–16S and vWF were as described elsewhere^{22,23}. Primers for RAG1, RAG2 and BRCA1 were designed on the basis of conserved regions of sequence alignments involving the few mammalian taxa available in GenBank. BRCA1 PCR primers were the following (numbers refer to the location of the 5' end of the primer sequence in human exon 11): BRCA1F126: 5'-GTTTCAAACCTTG-CATGTGGAGCC-3'; BRCA1R3012: 5'-GTTTGAAGCAGGGAAGCTCTTCATC-3'. PCR primers for RAG genes were the following (location of the 5' end of the primer sequence, in the single human exon of both genes, is indicated by the number included in the primer label): RAG1: RAG1F1705: 5'-GCTTTGATGGACATGGAAGAAGACAT-3'; RAG1R2864: 5'-GAGCCATCCTCTCAATAATTTCAGG-3'; RAG2: RAG2F220 5'-GATTCTCTGC-TA(CT)CT(TC)CCTCCTCT-3'; RAG2R995 5'-CCCATGTTGCTTCCAAACCATA-3'.

Phylogenetic analysis

Data sets for individual loci and combined data sets were all analysed with maximum parsimony, transversion parsimony, minimum evolution with maximum likelihood and logdet distances, and maximum likelihood. Individual data sets, for both 10 and 12 taxa, provided strong bootstrap support (that is, > 90% in parsimony analyses) for at least one clade (RAG2; 12 taxa) and as many as eight different clades (BRCA1; 12 taxa). In all parsimony analyses gaps were treated as missing data. Both parsimony and minimum evolution analyses employed tree bisection-reconnection branch swapping. In parsimony searches 10 random input orders were used. Parsimony and minimum evolution bootstrap analyses involved 500 replicates. Maximum likelihood trees were obtained with heuristic searches (10 random input orders) under the HKY-85 model of sequence evolution with a 2 to 1 transition to transversion ratio and an assumption of equal rates across sites for the mitochondrial, nuclear and mitochondrial-plus-nuclear data sets. For each of the resulting trees, we then obtained maximum likelihood estimates of the transition to transversion ratio (20 taxa: mitochondrial(m) = 3.29; 12 taxa: nuclear(n) = 2.47; mitochondrial-plus-nuclear (m+n) = 2.52; 10 taxa: m = 3.30; n = 2.45; m+n = 2.56) and the shape parameter for the gamma distribution (20 taxa: m = 0.25; 12 taxa: n = 0.77; m+n = 0.53; 10 taxa: m = 0.25; n = 61; m+n = 0.44). These values were used in subsequent maximum likelihood analyses, including bootstrap resampling (200 replications). Phylogenetic methods that were employed in outgroup sensitivity analyses included quartet puzzling, minimum evolution with maximum likelihood and logdet distances, and maximum parsimony. Kishino and Hasegawa²⁴ and Templeton²⁵ tests were used to evaluate *a priori* hypotheses. All analyses were performed using PAUP* 4.0b2 (ref. 7).

Received 20 October; accepted 15 November 1999.

- Novacek, M. J. In *Proceedings of the 5th International Bat Research Conference* (eds Wilson, D. E. & Gardner, A. L.) 317–330 (Texas Tech, Lubbock, 1980).
- Simmons, N. B. *Bat relationships and the origin of flight*. *Symp. Zool. Soc. London* **67**, 27–43 (1995).
- Hill, J. E. & Smith, J. D. In *Bats: A Natural History* (Heffers, England, 1984).
- Simmons, N. B. & Geisler, J. H. Phylogenetic relationships of *Icaronycteris*, *Archaeonycteris*, *Hassianycteris* and *Palaeochiropteryx* to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in Microchiroptera. *Bull. Am. Mus. Nat. Hist.* **235**, 1–182 (1998).
- Pumo, D. E. *et al.* Complete mitochondrial genome of a neotropical fruit bat, *Artibeus jamaicensis*, and a new hypothesis of relationships of bats to other Eutherian mammals. *J. Mol. Evol.* **47**, 709–717 (1998).
- Springer, M. S., Amrine, H. M., Burk, A. & Stanhope, M. J. Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with heterogeneous base composition. *Syst. Biol.* **48**, 65–75 (1999).
- Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. (Sinauer, Sunderland, MA, 1998).
- Cummings, M. P., Otto, S. P. & Wakeley, J. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.* **12**, 814–822 (1995).
- Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–514 (Sinauer, Sunderland, MA, 1996).
- Hutcheon, J. M., Kirsh, J. A. W. & Pettigrew, J. D. Base-compositional biases and the bat problem. III. The question of microchiropteran monophyly. *Phil. Trans. R. Soc. Lond. B* **353**, 607–617 (1998).
- Woodburne, M. O. & Swisher, C. C. in *Geochronology, Time Scales and Global Stratigraphic Correlation* (eds Beggren, W. A., Kent, D. V., Aubry, M. & Hardenbol, J.) *Special Publ. Soc. Sedim. Geol.* **54** 335–364 (Tulsa, Oklahoma, 1995).
- Altringham, J. D. *Bats: Biology and Behaviour*. (Oxford Univ. Press, New York, 1996).

13. Miller, G. S. The families and genera of bats. *Bull. US Natl Mus.* **57**, 1–282 (1907).
14. Novacek, M. J. & Wyss, A. R. Higher-level relationships of the recent eutherian orders: morphological evidence. *Cladistics* **2**, 257–287 (1986).
15. Beard, K. C. in *Primates and their relatives in phylogenetic perspective* (ed. MacPhee, R. D. E.) 63–90 (Plenum, New York, 1993).
16. Rayner, J. M. V. The cost of being a bat. *Nature* **350**, 383–384 (1991).
17. Gould, E. Evidence for echolocation in the Tenrecidae of Madagascar. *Proc. Am. Phil. Soc.* **109**, 352–360 (1965).
18. Tomasi, T. E. Echolocation by the short-tailed shrew *Blarina brevicauda*. *J. Mammal.* **60**, 751–759 (1979).
19. Forsman, K. A. & Malmquist, M. G. Evidence for echolocation in the common shrew, *Sorex araneus*. *J. Zool. Lond.* **216**, 655–662 (1988).
20. Speakman, J. R. & Racey, P. A. No cost of echolocation for bats in flight. *Nature* **350**, 421–423 (1991).
21. Kalko, E. K. V. Coupling of sound emission and wing-beat in naturally foraging European pipistrelle (*Microchiroptera: Vespertilionidae*). *Folia Zool.* **43**, 363–376 (1994).
22. Springer, M. S. *et al.* Endemic African mammals shake the phylogenetic tree. *Nature* **388**, 61–64 (1997).
23. Stanhope, M. J. *et al.* Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl Acad. Sci. USA* **95**, 9967–9972 (1998).
24. Kishino, H. & Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order within Hominoidea. *J. Mol. Evol.* **29**, 170–179.
25. Templeton, A. R. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37**, 221–244 (1983).
26. Lyons-Weiler, J., Hoelzer, G. A. & Tausch, R. J. Relative Apparent Synapomorphy Analysis (RASA) I: the statistical measurement of phylogenetic signal. *Mol. Biol. Evol.* **13**, 749–757 (1996).

Acknowledgements

This work was supported by grants from the Training and Mobility of Researchers programme of the European Commission to M.J.S., and the National Science Foundation to M.S.S.

Correspondence and requests for materials should be addressed to M.J.S. (e-mail: Michael_J_Stanhope@sbphrd.com) or M.S.S. (e-mail: mark.springer@ucr.edu).

Human cerebellar activity reflecting an acquired internal model of a new tool

Hiroshi Imamizu*, Satoru Miyauchi†, Tomoe Tamada*, Yuka Sasaki†‡, Ryosuke Takino§, Benno Pütz*‡, Toshinori Yoshioka* & Mitsuo Kawato*

* JST/ERATO Kawato Dynamic Brain Project, 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

† Communications Research Laboratory, 588-2, Iwaoka, Nishi-ku, Kobe, Hyogo 651-2401, Japan

§ Shiraume Gakuen College, 1-830, Ogawa-cho, Kodaira-shi, Tokyo 1878570, Japan

|| ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Theories of motor control postulate that the brain uses internal models of the body to control movements accurately. Internal models are neural representations of how, for instance, the arm would respond to a neural command, given its current position and velocity^{1–6}. Previous studies have shown that the cerebellar cortex can acquire internal models through motor learning^{7–11}. Because the human cerebellum is involved in higher cognitive function^{12–15} as well as in motor control, we propose a coherent computational theory in which the phylogenetically newer part of the cerebellum similarly acquires internal models of objects in the external world. While human subjects learned to use a new tool (a computer mouse with a novel rotational transformation), cere-

bellar activity was measured by functional magnetic resonance imaging. As predicted by our theory, two types of activity were observed. One was spread over wide areas of the cerebellum and was precisely proportional to the error signal that guides the acquisition of internal models during learning. The other was confined to the area near the posterior superior fissure and remained even after learning, when the error levels had been equalized, thus probably reflecting an acquired internal model of the new tool.

Most neuroimaging studies have found that the regional blood flow in the human cerebellum increases significantly at the beginning of learning for a new motor or cognitive task and decreases as the learning proceeds^{16–19}. These results are often interpreted as meaning that the cerebellum is involved only in the early phase of learning and is not a memory site, that is, it does not store internal models. Here we present a different interpretation (see also ref. 15) based on our computational theory and experimental results.

Previous cerebellar learning theories^{20–22} make no specific predictions about the activity of internal models (see Supplementary Information for details). We have proposed that multiple internal models exist and that they compete to learn new environments and tools²³. During the learning, all of these multiple internal models receive a copy of the error signal and only one or a few learn the new transformation, thereby reducing the error signal and localizing the new activity to a distinct region of the cerebellum. The two types of

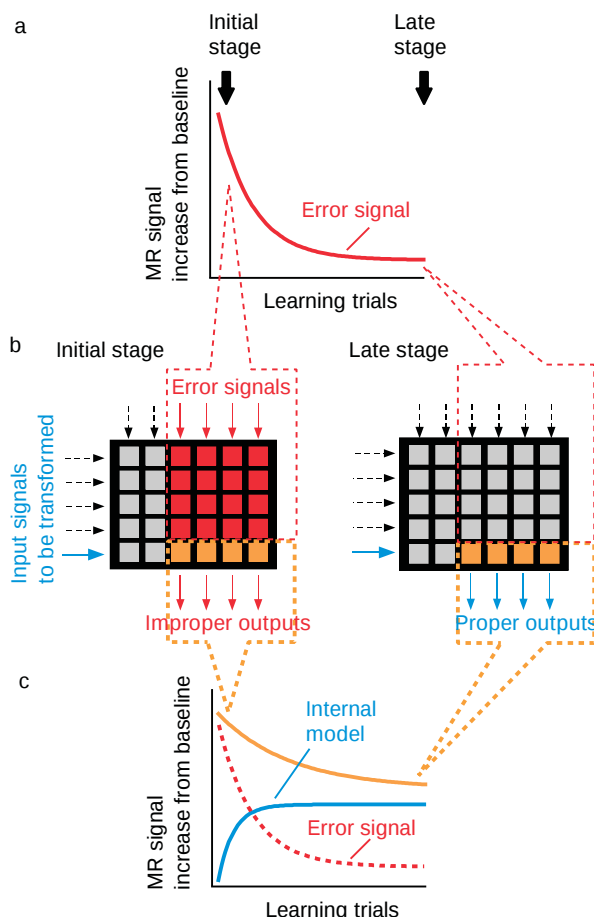


Figure 1 Changes in cerebellar activity predicted by the learning theory of internal models. **a**, Signal intensity change caused by error signals (red curve). **b**, Activity maps of the initial and late stages of learning. Each small square represents a unit of multiple internal models. Orange squares indicate regions where the internal model is acquired. **c**, Signal intensity change in the regions where the internal model is acquired. Orange curve indicates activity in the orange squares (the sum of activity reflecting error signals (red curve) and that reflecting the acquired internal model (cyan curve)).

‡ Present addresses: Massachusetts General Hospital NMR Center, 149, 13th Street, Charlestown, Massachusetts 02129, USA (Y.S.); Kernspintomographie, Max-Planck-Institut für Psychiatrie, Kraepelinstrasse 10, 80804 Muenchen, Germany (B.P.).

cerebellar activity representing the error signals and the internal model are predicted to occur in specific spatio-temporal patterns (Fig. 1). Large error signals are fed into broad regions for all possible internal model candidates at the initial learning stage (red and orange in Fig. 1b). Because the subject's performance improves and the error signals decrease with learning, activity reflecting the error signals also decreases with learning (Fig. 1a). This prediction agrees with previous imaging data. However, signals representing the acquired internal model must increase (cyan curve in Fig. 1c) and remain even at the late stages of learning only in the limited region (orange in Fig. 1b). This region contains a group of the most accurate internal models. Even in this region, total activity is expected to decrease (orange curve in Fig. 1c) because the observed activity is the sum of the error signal (broken red curve) and the internal model (cyan curve). However, significant activity must remain even after the learning is complete. In our interpretation, the activity reflecting the learned model is smeared by the strong activity reflecting the error signal and, therefore, was not detected in previous imaging studies. In our present study, however, we found internal model activity by equalizing the errors under the baseline condition to those under the test condition. Furthermore, the time course of the internal-model activity (cyan curve in Fig. 1c) was estimated by subtracting the error signal activity (red curve) from the total activity (orange curve).

The task for the subjects was to manipulate a computer mouse so that the corresponding cursor followed a randomly moving target on a screen (Fig. 2a; a tracking task). Seven subjects performed the task for eleven sessions (training sessions). We used functional magnetic resonance imaging (fMRI) to scan the cerebellum in the odd-numbered sessions. Each session lasted 9 min and 23 s and comprised eight alternating blocks of test and baseline periods. During the test periods, the cursor appeared in a position rotated 120° around the centre of the screen to necessitate subject learning (novel mouse); during the baseline periods, it was not rotated (ordinary mouse). In the first session (Fig. 2b), large regions near the posterior superior fissure in the lateral cerebellum were significantly more active during the test periods than the baseline periods (correlation coefficient (CC) > 0.3); in the last session (Fig. 2c), only restricted subregions were activated. We confirmed that this activity cannot be attributed to larger hand movements (Fig. 2d) or larger eye movements in the test than in the baseline periods by showing that hand-movement or eye-movement activity is different from the activity shown in Fig. 2b and c (see also Supplementary Information).

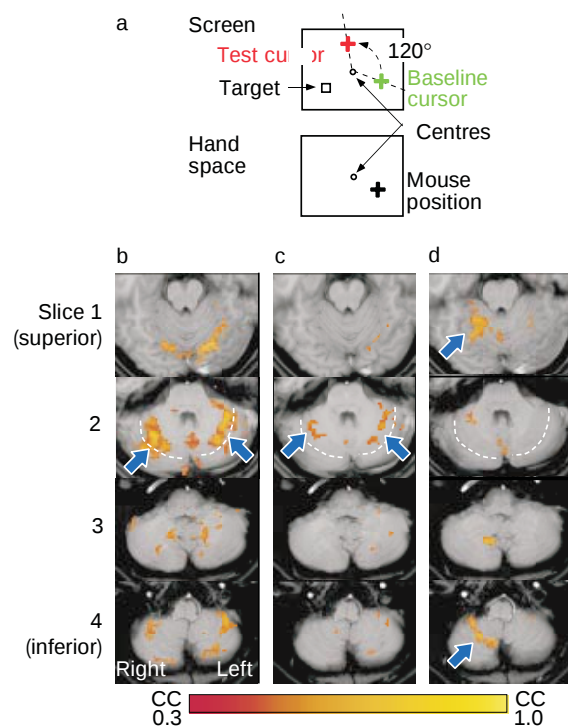


Figure 2 Visuomotor learning task and cerebellar activity. **a**, Rule for changing the relationship between mouse and cursor positions. A cross cursor coloured red or green appears on the screen in the test periods or baseline periods, respectively. The open circles indicate the centres of the screen and the hand space. **b**, Cerebellar activity of a typical subject in the first training session. The colour-coded regions were significantly activated in the test periods (CC > 0.3). The broken lines indicate the posterior superior fissure. **c**, Cerebellar activity of the same subject in the last training session. **d**, Regions that were more active during the tracking task (cursor position was not rotated) than during the task of pursuing the moving target with the eyes but without hand/mouse movements. These active regions are apparently different from those in **b** and **c** and are located in the superior anterior and inferior anterior parts of the cerebellum on the right side, ipsilateral to the hand that was performing the task.

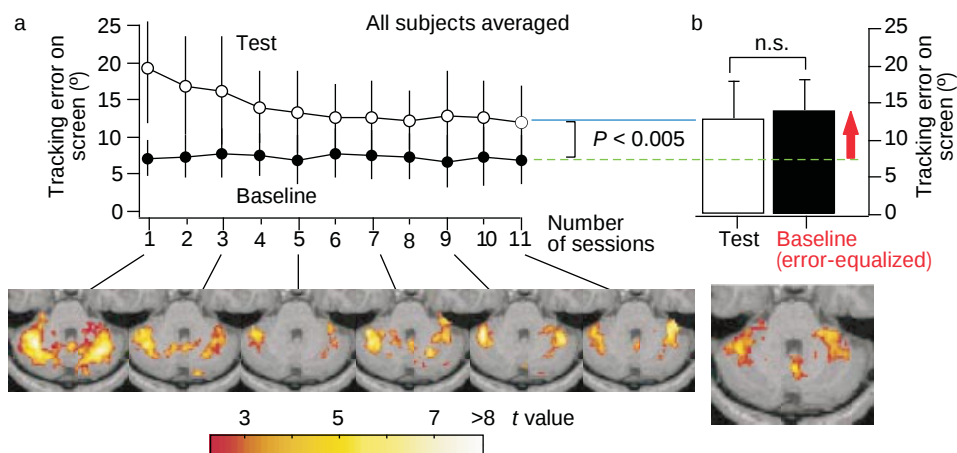


Figure 3 The two types of cerebellar activity. **a**, Activity that decreased with learning progress. Top, tracking error (mean $\pm$ s.d.) in the training sessions. Bottom, activation maps showing the regions that were significantly activated in the test periods in comparison to those in the baseline periods ($P < 0.05$ corrected). The slice position was

approximately the same as slice 2 in Fig. 2. **b**, The activity that remained when the tracking error was equalized. Top, tracking error (mean $\pm$ s.d.) in the error-equalized session. n.s.: not significant ($F(1, 6) = 4.52$). Bottom, activation map.

The subjects' performances were measured by tracking errors (the distance between the cursor and the target; see Methods). The errors in the test periods decreased significantly as the number of sessions increased, whereas the errors in the baseline periods were constant (upper panel in Fig. 3a). A repeated-measures analysis of variance (ANOVA) on the errors indicated a significant effect of the sessions in the test periods ($F(10, 60) = 10.60, P < 0.001$) but no significant effect in the baseline periods ($F(10, 60) = 0.67$). Activation maps ($P < 0.05$, corrected for multiple comparisons) derived from data across all subjects (Fig. 3a) indicated that the activity in the lateral cerebellum became smaller as learning progressed (see Methods).

The learning during the test periods was sufficient for there to be no significant difference in the tracking error between any pairs of the last three sessions according to the *post hoc* test (at $P < 0.001$ level by Tukey's honestly significant difference method). Thus, the cerebellar activation observed in the late stage of learning should include the activity of the acquired internal model. However, we cannot conclude that the activation solely reflects the internal model because the test error was close to, but significantly larger

than, the baseline error (for example, $F(1, 6) = 28.52, P < 0.005$ on the error in the last training session across all subjects). That is, the activation may partly reflect error signals.

To evaluate this possibility, all of the subjects underwent an 'error-equalized' experiment: the target velocity in the baseline periods was increased so that the baseline error was equal to the test error. Here, we used the linear relationship between error and target velocity (see Methods). As result, there was no significant difference between the test and the baseline errors (Fig. 3b, top). However, regions near the posterior superior fissure were significantly more active during the test periods than during the baseline periods (Fig. 3b, bottom). This activity cannot be related to the tracking error. Moreover, the amount of mouse movement (measured by the cursor trajectory length) and the target velocity in the baseline periods were significantly larger than those in the test periods ($F(1, 6) = 38.16, P < 0.001$, on average 2.38-fold, and $F(1, 6) = 156.63, P < 0.001$, on average 2.71-fold, respectively). All of the subjects reported that more effort and attention were needed in the baseline periods than in the test periods. Thus, the significant activity increase in the test period cannot be attributed to the mouse/hand movements, the visual target velocity, attention or effort. The most plausible explanation is that the remaining activity in Fig. 3b reflects the acquired internal models, whereas the decrease in activity as learning progresses (Fig. 3a) may largely reflect the error signals.

To strengthen the above conclusion quantitatively, we examined the time courses of gross signal intensity during all sessions averaged over two regions of interest. First, the error-related region (red and orange in Fig. 4b) was defined as voxels whose signal intensity during all the training sessions was significantly and positively correlated with the tracking error (that is, the estimated regression coefficient was significantly larger than zero, $t(5276) > 2.33, P < 0.05$ corrected). Second, the internal-model-related region (blue and orange in Fig. 4b) was defined as voxels whose signal intensity during the error-equalized session was significantly and positively correlated with the explanatory variable which takes 1 in the test period scans and 0 in the baseline period scans, and thus represents internal-model activity ($t(874) > 2.33, P < 0.05$ corrected). Orange voxels were correlated with both the tracking error and the internal model activity. The error-related region is widely spread over the lateral cerebellum, whereas the internal model seems to be acquired only in the restricted subregions.

The relative activity in the red and orange regions (per cent of mean signal increase from the baseline periods, all subjects averaged) decreased as the session number increased (red curve in Fig. 4a) and was highly correlated with the tracking error (per cent of increase from the baseline) indicated by the black curve ($r^2 = 0.82$ for all sessions, $F(1, 5) = 22.87, P < 0.005$). In contrast, the activity in the blue and orange regions did not markedly decrease (orange curve in Fig. 4c), and its correlation with the error was low ($r^2 = 0.25, F(1, 5) = 1.68$). The signal increase in the blue and orange regions was significantly larger than that in the red and orange regions ($F(1, 6) = 7.38, P < 0.05$). These data indicate that the activity in the blue and orange regions may include components that cannot be explained solely by the error. The cyan curve in Fig. 4c shows the subtraction of the red curve from the orange curve, and represents the internal model activity according to our theory. This activity increased at the initial phase of learning and remained high even in the late learning stage where the error was equalized (Fig. 4c, right).

The acquired internal models in these experiments are expected to represent the altered relationship between the cursor movement and the mouse movement (forward and/or inverse kinematics models of the novel tool). We believe that these internal models were stored in different regions from those for an ordinary mouse, as no significant activity was observed near the posterior superior fissure when the subjects used an ordinary mouse in the test periods

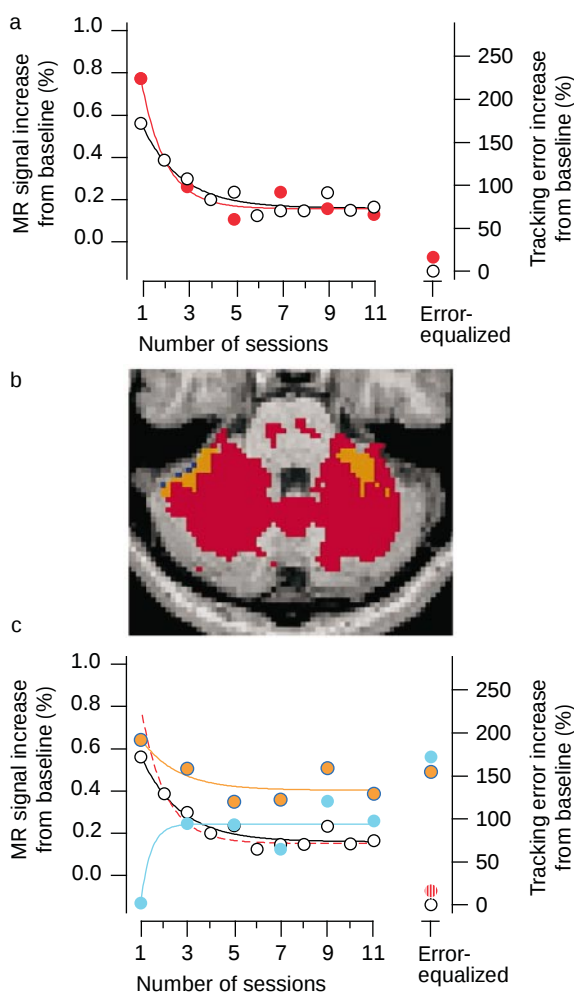


Figure 4 Cerebellar activity related to error signals and activity related to the acquired internal model. **a**, Activity change (red circles) in the red and orange regions in **b** in comparison to the decrease in tracking error (black circles). Each curve indicates the exponential function fitted to the circles. **b**, Activity in the red and orange regions was significantly ($P < 0.05$) correlated with the tracking error in the training sessions. Activity in the blue and orange regions was correlated with a step function representing the internal-model activity in the error-equalized session. **c**, Activity change in the blue and orange regions (blue circle filled by orange). The cyan circle indicates the subtraction of the activity in the red region from that in the blue and orange regions. The black circles and solid curve indicate the tracking error increase (as in **a**). The red broken curve is a duplication of the red curve in **a**.

and pursued the moving target with the eyes but without hand/mouse movements in the baseline periods (see Fig. 2d and Supplementary Information). According to an electrophysiological study in monkeys²⁴, regions near this fissure receive parallel fibre inputs from the premotor and parietal association cortex, and are thus suitable to represent kinematic models of tools. The bilateral activity (see ref. 25 for related bilateral activity) may indicate that activated regions acquire internal models for cognitive function independent of the ipsilateral correspondence between the motor apparatus and the cerebellum. Whereas previous neurophysiological experiments indicated that internal models for the motor apparatus are present in phylogenetically older parts of the cerebellum (such as the ventral paraflocculus, vermis and intermediate parts)^{8–11}, internal models of objects and tools in the external world seem to reside in newer parts. We further speculate that the cerebellum assists information processing in cerebral areas by providing general internal models of extended controlled objects in the external world such as concepts, symbols and languages. □

Methods

Task

The subjects moved a computer mouse (PocketEgg, Elecom) using the right hand while lying in an MRI scanner. Head movements were restrained by a bite bar. They used a tilted mirror to view a rear-projection screen outside the scanner. A colour projector (VPH-12720 LCD; Sony) controlled by a computer (PC-9821 AP2; NEC) displayed the target and the cursor on the screen. During the tracking task, a small white square target was presented on a dark background. The *x* and *y* components of the target path were each sums of sinusoids whose amplitude and frequency were pseudorandomly determined. The subjects moved a small cross-hair cursor on the screen with the mouse. The cursor position was sampled at 60 Hz. The distance between the cursor and the target at each sampling point was accumulated over 4.4 s (tracking error).

Subjects

Ten neurologically normal subjects (20–34 years of age; five females and five males) participated in the experiments. Each participant gave informed written consent. Seven of the subjects (five right-handed and two left-handed²⁶) underwent the training sessions and the error-equalized experiment. The other three subjects (two right-handed and one left-handed) underwent the training sessions and control experiments. In the control experiments, we confirmed that the activity observed in the above seven subjects could not be attributed to differences in hand or eye movements (see Fig. 2d and Supplementary Information).

MRI acquisition

A 1.5-T MRI scanner (Magnetom Vision; Siemens) was used to obtain blood oxygen level-dependent contrast functional images. Images weighted with the apparent transverse relaxation time (*T*₂) were obtained with an echo-planar imaging sequence (repetition time (TR) 4.4 s, echo time (TE) 66 ms, flip angle (FA) 90°, field of view (FoV) 240 mm × 240 mm, matrix size 128 × 128). We selected ten axial slices (thickness 7 mm, slice gap 0.21 mm) encompassing the cerebellum. We scanned 128 functional images for each slice during one session. Anatomical images for these slices were obtained with a *T*₁ weighted sequence (TR 350 ms, TE 6 ms, FA 90°, FoV 240 mm × 240 mm, matrix size 256 × 256).

MRI analysis

Motion artefacts in all functional images were removed by using Automated Image Registration (AIR) version 3.0 (ref. 27). We used two approaches to analyse the functional images: a correlation analysis on a pixel-by-pixel basis²⁸ and an analysis based on the general linear model as implemented in SPM99b (Wellcome Department of Cognitive Neurology). Details of the correlation analysis have been reported in ref. 29. To analyse group data, functional images of each subject's cerebellum were stereotactically transformed to a standard template in SPM, and were smoothed with a gaussian kernel 4 mm full width at half maximum (FWHM). In the activation analyses shown in Fig. 3, condition-specific effects were estimated with the linear model with a boxcar wave form. Areas of significant change in brain activity were specified by linear contrasts of the condition-specific effects and determined using the *t*-statistics (SPM{*t*}). Results were thresholded at *t*-value 2.33. In assessing the statistical significance of each cluster, we corrected for multiple comparisons based on random gaussian field theory in terms of spatial extent and/or peak height (*P* < 0.05). Voxel time series were temporally smoothed with a gaussian filter (FWHM of 4 s). We used the effective degree of freedom adjusted for analysis of fMRI time-series³⁰. In the activation analyses shown in Fig. 4b, the explanatory variable of main interest was either the tracking error or the internal model activity.

Equalization of the tracking error according to the relationship between the error and the target velocity

The subjects performed the tracking task under various target velocities for about 15 min. The cursor position was not rotated during this task. The averaged target velocity

(21.30° s⁻¹) used in the training sessions was multiplied by a value ranging from 1.0 to 5.0 at intervals of 0.2. Therefore, the target moved at 21 different averaged velocities in random order. Then, the relationship was estimated linearly by the least-squares method. The effect of the velocity on the error was significant (*r*² > 0.70, *F*(1, 19) > 45.27, *P* < 0.0001) for each subject. When the cerebellar activity was scanned, the target velocity was increased in the baseline period using this relationship, so that the baseline error was equal to the mean error in the preceding test period.

Received 12 July; accepted 14 October 1999.

- Kawato, M., Furukawa, K. & Suzuki, R. A hierarchical neural-network model for control and learning of voluntary movement. *Biol. Cybern.* **57**, 169–185 (1987).
- Lackner, J. R. & Dizio, P. Rapid adaptation to Coriolis force perturbations of arm trajectory. *J. Neurophysiol.* **72**, 299–313 (1994).
- Shadmehr, R. & Mussa-Ivaldi, F. A. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
- Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. An internal model for sensorimotor integration. *Science* **269**, 1880–1882 (1995).
- Imamizu, H., Uno, Y. & Kawato, M. Internal representations of the motor apparatus: implications from generalization in visuomotor learning. *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 1174–1198 (1995).
- Gomi, H. & Kawato, M. Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement. *Science* **272**, 117–120 (1996).
- Kawato, M. & Gomi, H. A computational model of four regions of the cerebellum based on feedback-error learning. *Biol. Cybern.* **68**, 95–103 (1992).
- Shidara, M., Kawano, K., Gomi, H. & Kawato, M. Inverse-dynamics model of eye movement control by Purkinje cells in the cerebellum. *Nature* **365**, 50–52 (1993).
- Gomi, H. *et al.* Temporal firing patterns of Purkinje cells in the cerebellar ventral paraflocculus during ocular following responses in monkeys I. Simple spikes. *J. Neurophysiol.* **80**, 818–831 (1998).
- Kobayashi, Y. *et al.* Temporal firing patterns of Purkinje cells in the cerebellar ventral paraflocculus during ocular following responses in monkeys II. Complex spikes. *J. Neurophysiol.* **80**, 832–848 (1998).
- Kitazawa, S., Kimura, T. & Yin, P. B. Cerebellar complex spikes encode both destinations and errors in arm movements. *Nature* **392**, 494–497 (1998).
- Raichle, M. E. *et al.* Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb. Cortex* **4**, 8–26 (1994).
- Kim, S. G., Ugurbil, K. & Strick, P. L. Activation of a cerebellar output nucleus during cognitive processing. *Science* **265**, 949–951 (1994).
- Allen, G., Buxton, R. B., Wong, E. C. & Courchesne, E. Attentional activation of the cerebellum independent of motor involvement. *Science* **275**, 1940–1943 (1997).
- Thach, W. T. On the specific role of the cerebellum in motor learning and cognition: Clues from PET activation and lesion studies in man. *Behav. Brain Sci.* **19**, 411–431 (1996).
- Friston, K. J., Frith, C. D., Passingham, R. E., Liddle, P. F. & Frackowiak, R. S. Motor practice and neurophysiological adaptation in the cerebellum: a positron tomography study. *Proc. R. Soc. Lond. B Biol. Sci.* **248**, 223–228 (1992).
- Grafton, S. T., Woods, R. P. & Tyszka, M. Functional imaging of procedural motor learning: relating cerebral blood flow with individual subject performance. *Hum. Brain Mapp.* **1**, 221–234 (1994).
- Seitz, R. J. *et al.* Successive roles of the cerebellum and premotor cortices in trajectory learning. *NeuroReport* **5**, 2541–2544 (1994).
- Flament, D., Ellermann, J. M., Kim, S. G., Ugurbil, K. & Ebner, T. J. Functional magnetic resonance imaging of cerebellar activation during the learning of a visuomotor dissociation task. *Hum. Brain Mapp.* **4**, 210–226 (1996).
- Marr, D. A theory of cerebellar cortex. *J. Physiol. (Lond.)* **202**, 437–470 (1969).
- Albus, J. S. A theory of cerebellar function. *Math. Biosci.* **10**, 25–61 (1971).
- Ito, M. Cerebellar control of the vestibulo-ocular reflex—around the flocculus hypothesis. *Annu. Rev. Neurosci.* **5**, 275–296 (1982).
- Wolpert, D. M. & Kawato, M. Multiple paired forward and inverse models for motor control. *Neural Netw.* **11**, 1317–1329 (1998).
- Sasaki, K. *et al.* Mossy fibre and climbing fibre responses produced in the cerebellar cortex by stimulation of the cerebral cortex in monkeys. *Exp. Brain Res.* **29**, 419–428 (1977).
- Roland, P. E., Eriksson, L., Widen, L. & Stone-Elander, S. Changes in regional cerebral oxidative metabolism induced by tactile learning and recognition in man. *Eur. J. Neurosci.* **1**, 3–18 (1988).
- Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
- Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R. & Mazziotta, J. C. Automated image registration: I. General methods and intrasubject, intramodality validation. *J. Comput. Assist. Tomogr.* **22**, 139–152 (1998).
- Bandettini, P. A., Jesmanowicz, A., Wong, E. C. & Hyde, J. S. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn. Reson. Med.* **30**, 161–173 (1993).
- Tamada, T., Miyachi, S., Imamizu, H., Yoshioka, T. & Kawato, M. Cerebro-cerebellar functional connectivity revealed by the laterality index in tool-use learning. *NeuroReport* **10**, 325–331 (1999).
- Worsley, K. J. & Friston, K. J. Analysis of fMRI time-series revisited—again. *Neuroimage* **2**, 173–181 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial offices of Nature.

Acknowledgements

We thank D. Wolpert, C. Miall, M. Honda, K. Sakai and S. Kitazawa for comments on the manuscript. Supported by Human Frontier Science Projects, and Special Coordination Fund to Brain Science.

Correspondence and requests for materials should be addressed to H.I. (e-mail: imamizu@erato.atr.co.jp).

A constitutively open potassium channel formed by KCNQ1 and KCNE3

Björn C. Schroeder^{*,†}, Siegfried Waldegger^{*,†}, Susanne Fehr^{*}, Markus Bleich[‡], Richard Warth[‡], Rainer Greger[‡] & Thomas J. Jentsch^{*}

^{*} Zentrum für Molekulare Neurobiologie Hamburg (ZMNH), Hamburg University, Martinistrasse 85, D-20246 Hamburg, Germany

[‡] Physiologisches Institut der Albert-Ludwigs-Universität Freiburg, Hermann-Herder-Strasse 7, D-79104 Freiburg, Germany

[†] These authors contributed equally to this work

Mutations in all four known KCNQ potassium channel α -subunit genes lead to human diseases^{1–6}. KCNQ1 (KvLQT1)¹ interacts with the β -subunit KCNE1 (IsK, minK)⁷ to form the slow, depolarization-activated potassium current I_{Ks} ^{8,9} that is affected in some forms of cardiac arrhythmia. Here we show that the novel β -subunit KCNE3 markedly changes KCNQ1 properties to yield currents that are nearly instantaneous and depend linearly on voltage. It also suppresses the currents of KCNQ4 and HERG potassium channels. In the intestine, KCNQ1 and KCNE3 messenger RNAs colocalized in crypt cells. This localization and the pharmacology, voltage-dependence and stimulation by cyclic AMP of KCNQ1/KCNE3 currents indicate that these proteins may assemble to form the potassium channel that is important for cyclic AMP-stimulated intestinal chloride secretion and that is involved in secretory diarrhoea and cystic fibrosis.

KCNQ proteins have six transmembrane domains and probably function as tetramers. In the heart and in the cochlea, KCNQ1 associates with KCNE1, a small protein which spans the membrane only once⁷. Dominant mutations in either KCNQ1 (refs 10, 11) or KCNE1 (ref. 12) prolong cardiac action potentials in long-QT syndrome. A total loss of function additionally causes deafness² by impairing potassium secretion across the stria vascularis of the cochlea. The positive apical membrane potential of these cells¹³ allows KCNQ1/KCNE1 channels to be open. KCNQ1 is also expressed in several other tissues¹⁰, but its role in these organs is unclear. Pharmacological data have indicated that KCNQ1 may be important for chloride secretion in the colon¹⁴. However, the negative membrane voltage of intestinal epithelial cells would prevent KCNQ1 or KCNQ1/KCNE1 from being open. This suggests that there may be another subunit.

We cloned the β -subunit KCNE3 (ref. 15) by homology to KCNE1. It is a 103-amino-acid protein with a relative molecular mass of $\sim 12,000$ ($M_r \approx 12K$) that displays roughly 35% identity to other KCNE proteins within the single transmembrane domain and a following short stretch (Fig. 1). Northern analysis revealed a prominent band in the kidney, moderate expression in the small intestine and weaker bands in most other tissues including colon and heart (see Supplementary Information).

We co-expressed KCNE3 with KCNQ1 through KCNQ4. KCNQ1

gave potassium currents that slowly activated at potentials more positive than -40 mV (Fig. 2a). Co-expression with KCNE1 enhanced the current amplitudes, slowed channel activation and shifted the voltage dependence to more positive potentials^{8,9} (Fig. 2b). Co-expression with KCNE3 caused a marked change in the channel properties (Fig. 2c, d). Currents now had a linear I/V relationship, were also present at negative voltages and were nearly instantaneous. There was some time-dependent gating at positive voltages. Ion substitution experiments showed that the KCNQ1/KCNE3 channel is highly selective for potassium (Fig. 2e). Co-expression of KCNQ1 with KCNE1 and KCNE3 gave currents that could not be distinguished from a linear superposition of KCNQ1/KCNE1 and KCNQ1/KCNE3 channels (data not shown). KCNQ1 currents are specifically inhibited by chromanol 293B, and the sensitivity of KCNQ1/KCNE1 heteromers is increased¹⁶ by a factor of ~ 10 . Chromanol 293B at $10 \mu M$ inhibited KCNQ1 channels by $\sim 20\%$, and this inhibition was increased to $\sim 85\%$ when KCNE3 was co-expressed (Fig. 2f). KCNQ1/KCNE3 channels were inhibited by chromanol 293B with an inhibitory constant $K_i \approx 3 \mu M$ (Fig. 2g). KCNE3 also increased the sensitivity to clotrimazole, an inhibitor of distinct calcium- and cAMP-activated potassium channels^{17,18} (Fig. 2f). Barium (5 mM) inhibited KCNQ1/KCNE3 currents by $\sim 70\%$ (data not shown). Like KCNQ1 (ref. 19) and KCNQ2/KCNQ3 heteromers²⁰, KCNQ1/KCNE3 currents were enhanced by increasing intracellular cAMP (Fig. 2h). To confirm that KCNQ1 and KCNE3 interact physically, we transfected epitope-tagged versions of both subunits into mammalian cells. Immunofluorescence showed that KCNQ1 resided in the plasma membrane irrespective of whether it was co-expressed with KCNE3. However, KCNE3 was distributed in small vesicles in the cytoplasm when expressed alone (Fig. 3a), but was transported to the plasma membrane where it colocalized with KCNQ1 when they were expressed together (Fig. 3b, c). Also, in *Xenopus* oocytes, co-injection with KCNQ1 greatly increased surface expression²¹ of KCNE3 (Fig. 3d).

In contrast, currents from KCNQ2/KCNQ3 channels (which are mutated in neonatal epilepsy²⁰) were not significantly affected by the same level of KCNE3 expression (Fig. 4a, b). Currents from KCNQ4, a channel that is mutated in dominant deafness⁶, were largely suppressed by KCNE3 (Fig. 4c, d). Currents of the HERG potassium channel are also modulated by KCNE1 (ref. 22) and KCNE2 (ref. 15). Co-expression of KCNE3 strongly inhibited HERG currents (Fig. 4e, f).

As KCNE3 expression in heart is weak compared with those of HERG²³ and KCNE1 (ref. 10), it is unclear whether the effect on HERG is physiologically important. *In situ* hybridization of cochlear sections with a KCNE3 probe revealed no specific signal (J. P. Hardelin and C. Petit, personal communication), indicating that it may not suppress KCNQ4 currents in sensory outer hair cells⁶.

The interaction with KCNQ1, however, is highly significant. KCNE3 and KCNQ1 are expressed together in various tissues, including the small intestine, colon and kidney^{10,19}. The KCNQ1/KCNE3 channel

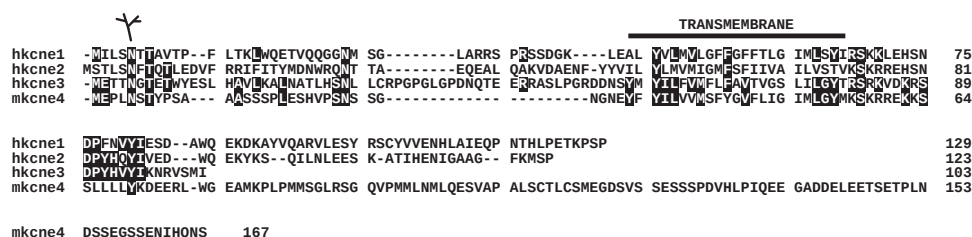


Figure 1 Primary structure of KCNE3. Amino-acid sequence comparison of human KCNE3 (third line) with other KCNE proteins. Black background indicates residues that are identical in KCNE3 and at least one other KCNE protein. The bar indicates the predicted transmembrane domain. There is a conserved consensus site for *N*-glycosylation near the

amino terminus, as indicated by the branched symbol. It is likely that all these proteins have an extracellular N terminus, as also supported by the experiment shown in Fig. 3d. Accession numbers are M26685, KCNE1; AF071002, KCNE2; AF076531, KCNE3; AF076533, KCNE4 (ref. 17).

is open at resting potentials and is therefore suitable for mediating transepithelial transport. Several reports stress the importance of a basolateral potassium conductance in cAMP-stimulated secretion of chloride by colonic crypt cells^{17,18,24–28}. This channel is thought to recycle potassium that is transported into the cell by the basolateral NaK2Cl cotransporter. It thereby stimulates basolateral uptake of chloride and hyperpolarizes the cell. Both effects increase the driving force for the apical exit of chloride ions through the CFTR chloride channel. The basolateral cAMP-dependent potassium conductance was sensitive^{14,27,28} to chromanol 293B, a specific inhibitor of KCNQ1/KCNE1 (ref. 16). However, the voltage dependence of KCNQ1 (and KCNQ1/KCNE1) channels predicts that they would be closed at resting voltages. It also differs from the voltage dependence of chromanol-sensitive potassium currents in native intestinal crypt cells²⁹. The chromanol-sensitive component of potassium currents is instantaneous, has a linear *I/V* relationship and is enhanced by intracellular cAMP (Fig. 5). It is also inhibited by clotrimazole (data not shown). A similar linear *I/V* relationship was

observed for a cAMP-activated, clotrimazole-sensitive, basolateral potassium conductance in apically permeabilized T84 colon carcinoma cell monolayers¹⁸.

Our work indicates that KCNQ1/KCNE3 may be the molecular basis of this colonic channel. KCNE3 renders KCNQ1 a constitutively open channel with a linear *I/V* relationship. KCNQ1/KCNE3 channels share many other features with the native current of colonic crypt cells: both are stimulated by cAMP^{17,18,25–28} and are sensitive to micromolar amounts of chromanol 293B^{24–28}. In the human colon, chromanol 293B inhibited cAMP-stimulated secretion of chloride with a similar inhibitory constant ($K_i \approx 5 \mu\text{M}$)²⁶ as that with which it inhibited human KCNQ1/KCNE3 in *Xenopus* oocytes ($\sim 3 \mu\text{M}$; Fig. 2g). Both currents are also inhibited by clotrimazole, which blocks both cAMP- and calcium-stimulated

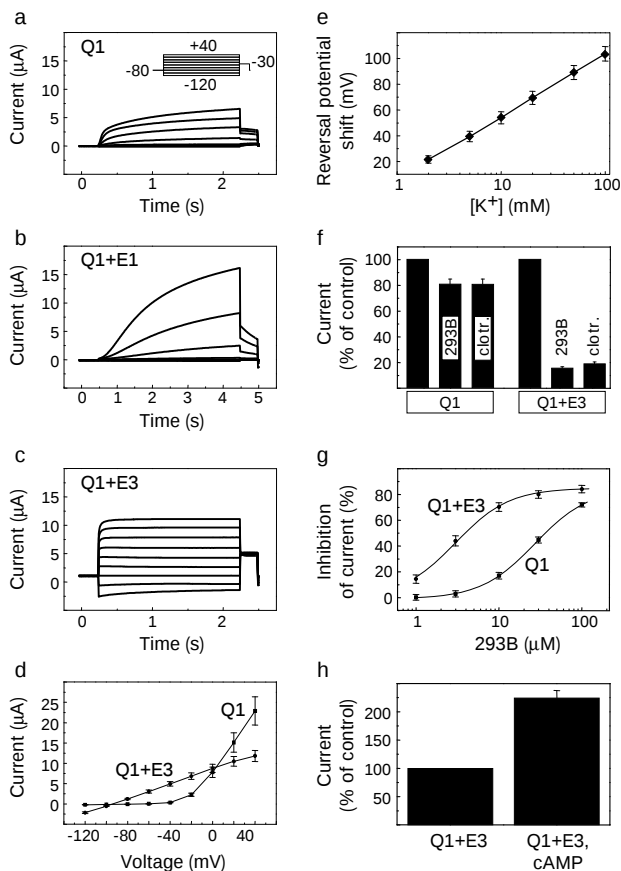


Figure 2 Modification of KCNQ1 currents by KCNE1 and KCNE3. Two-electrode voltage clamp traces of *Xenopus* oocytes injected with cRNA encoding KCNQ1 (**a**), KCNQ1 and KCNE1 (**b**) and KCNQ1 and KCNE3 (**c**). Voltage was clamped between -120 and +40 mV (**a**, inset). Longer pulses were used in **b**, **d**. **d**, Averaged current/voltage relationships for KCNQ1 (squares) and KCNQ1/KCNE3 (circles). Currents obtained after 2 s at the respective voltages were averaged from 13 oocytes. **e–h**, Characterization of KCNQ1/KCNE3 currents. **e**, Shift of reversal potential as a function of extracellular K⁺ concentration (53 mV per decade). Results are from 15 oocytes from two different batches. **f**, Effect of chromanol 293B (10 μM) and clotrimazole (clotr.; 20 μM) on KCNQ1/KCNE3 and KCNQ1 (averaged from ≥5 oocytes). **g**, Dose-response curve for chromanol 293B of KCNQ1 (squares) and KCNQ1/KCNE3 (circles) channels. Currents were measured after 2 s at clamping to +40 mV and averaged from six oocytes. This yields $K_i \approx 27 \mu\text{M}$ for KCNQ1 and $K_i \approx 3 \mu\text{M}$ for KCNQ1/KCNE3. **h**, Effect of raising intracellular cAMP on KCNQ1/KCNE3 currents. Results are from 10 oocytes from two different batches. Error bars indicate s.e.m. throughout.

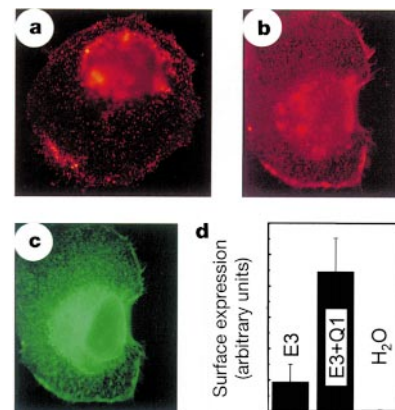


Figure 3 Stimulation of KCNE3 surface expression by KCNQ1. **a**, KCNE3 is localized in numerous intracellular vesicles in a CHO cell transfected with epitope-tagged KCNE3. **b**, Surface staining for KCNE3 in a CHO cell cotransfected with KCNQ1. **c**, The same cell stained for KCNQ1. As is common with overexpression of membrane proteins, there is also labelling of intracellular structures which may correspond to the endoplasmic reticulum and Golgi. **d**, Surface expression of epitope-tagged KCNE3 in *Xenopus* oocytes using a luminescence detection method²¹. The surface expression of KCNE3 in the absence of exogenous KCNQ1 may be due to xKCNQ1 endogenous to *Xenopus* oocytes⁹. Mean values from 15 oocytes are shown, with error bars indicating s.e.m.

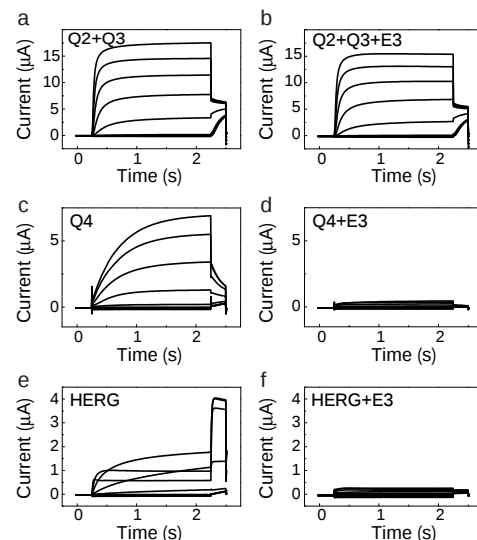


Figure 4 Effect of KCNE3 on other channels. **a**, **b**, Effects on KCNQ2/KCNQ3 heteromers; **c**, **d**, effects on KCNQ4; **e**, **f**, effects on HERG. Two-electrode voltage clamp measurements of *Xenopus* oocytes, using the clamp protocol of inset in Fig. 2a. Currents were averaged from more than seven oocytes. KCNE3 was co-injected in **b**, **d**, **f**, but was lacking in **a**, **c**, **e**.

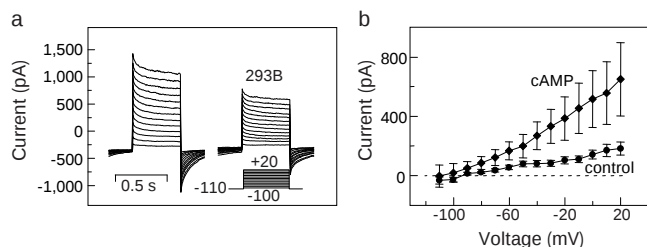


Figure 5 Chromanol 293B-sensitive currents in rat colonic crypt cells. **a**, Whole-cell voltage-clamp traces before and after addition of 10 μM chromanol 293B. From a holding potential of -110 mV, the voltage was clamped for 500 ms to values between -100 and +20 mV. The channels were activated by raising intracellular cAMP using 10 μM forskolin and 100 μM 8-CPT-cAMP, and were measured in the absence of Cl⁻ and the presence of 10 nM charybdotoxin to inhibit Cl⁻ channels and Ca²⁺-activated K⁺ channels, respectively. Chromanol-sensitive currents showed no time-dependent relaxations even with pulses of 2 s (not shown). **b**, Current/voltage relationship of chromanol-sensitive current before and after adding forskolin/cAMP.

intestinal chloride secretion^{17,18} and which can block cholera-toxin-induced diarrhoea in animals¹⁸.

In situ hybridization revealed both KCNQ1 and KCNE3 in crypt cells of the small intestine and the colon (Fig. 6a–f). Consistent with the finding that crypts are the predominant site of cAMP-stimulated chloride and fluid secretion, expression of both KCNQ1 and KCNE3 decreased along the crypt–villus axis and paralleled that of the CFTR chloride channel³⁰. Importantly, cAMP-activated secretion of chloride ions in the jejunum is also inhibited by chromanol 293B (data not shown). Northern analysis revealed that both KCNQ1 and KCNE3, but not KCNE1, are expressed in T84 colonic epithelial cells (Fig. 6g). These cells are a useful model system to study the importance of basolateral potassium channels in chloride secretion^{17,18,25} and express the cAMP-activated channel that is sensitive to chromanol and clotrimazole. By contrast, HEK293 kidney cells do not express these proteins.

Our work shows an unexpected effect on the gating of KCNQ1 of

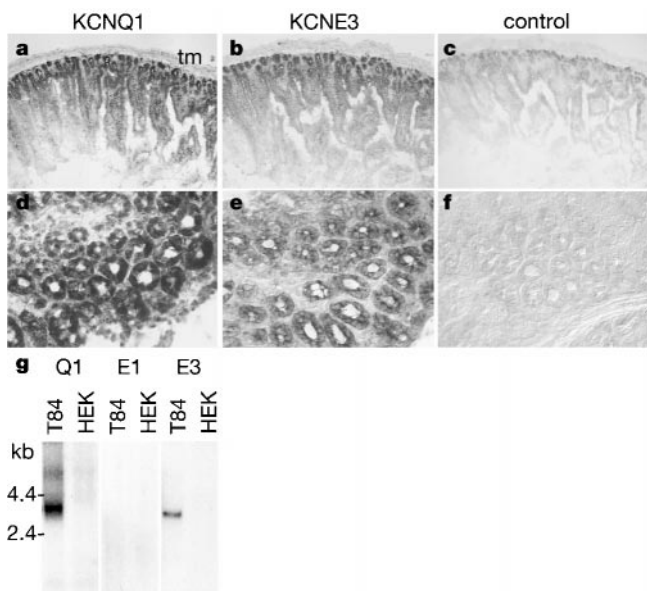


Figure 6 Localization of KCNQ1 and KCNE3 in the mouse small intestine and colon by *in situ* hybridization. **a–c**, Small intestine sections hybridized with a KCNQ1 antisense probe (**a**), a KCNE3 antisense probe (**b**) and a KCNQ1 sense probe as a control (**c**). tm, tunica muscularis. **d–f**, Colon crypts hybridized to KCNQ1 (**d**) and KCNE3 (**e**) antisense probes, and a KCNE3 sense probe (**f**) as negative control. **g**, Northern analysis of T84 human colon epithelial cells and HEK293 human embryonic kidney cells for expression of KCNQ1, KCNE1 and KCNE3. 25 μg total RNA was loaded per lane.

a protein with a single transmembrane domain. It differs fundamentally from the effect of the structurally related KCNE1 subunit, promising interesting insight into mechanism of channel gating. Furthermore, our results indicate that KCNQ1 and KCNE3 may form the cAMP-activated potassium channel that participates in cAMP-stimulated intestinal secretion of chloride ions, which is reduced in cystic fibrosis and pathologically stimulated in cholera and other forms of secretory diarrhoea. The rather broad and partially overlapping tissue distributions of KCNQ1 and KCNE3 indicate that KCNQ1/KCNE3 channels may exert similar functions in other tissues as well.

Methods

Cloning and functional expression of KCNE3

KCNE3 was cloned by PCR on a clone from human colon complementary DNA (accession no. AA133012) which we identified by a BLAST search of the dBEST database using the KCNE2 sequence that we had identified by homology to KCNE1. The KCNE3 sequence has been deposited into the database³¹, but no effects on KCNQ and HERG channels had been detected. The KCNE3 cDNA was inserted into vector pTLN⁶. From this clone (1 ng per oocyte) and from KCNQ and HERG clones (10 ng per oocyte, respectively) was synthesized and expressed and measured in *Xenopus* oocytes as described⁶. In most cases, KCNE3 injection by itself gave no currents. Sometimes there were small currents that may have been due to an interaction with the xKvLQT1 protein that is endogenous to *Xenopus* oocytes⁹.

Measurements were performed in ND96 (96 mM NaCl, 2 mM KCl, 1.8 mM CaCl₂, 1 mM MgCl₂, 5 mM HEPES pH 7.4). In Fig. 2d, KCl was exchanged for equivalent amounts of NaCl. Chromanol 293B³⁴ was added from a 10 mM stock, and clotrimazole from a 100 mM stock (both in DMSO). Intracellular cAMP was raised in *Xenopus* oocytes by applying 1 mM IBMX and 10 μM forskolin ≥ 10 min.

Localization of KCNE3 and KCNQ1

In situ hybridization was performed on cryosections of mouse small intestine and colon using mouse KCNQ1 and KCNE3 antisense and sense cRNAs in the DIG labelling system (Roche). For immunodetection, KCNQ1 and KCNE3 were tagged at the amino terminus with the FLAG- and HA-epitope, respectively, subcloned into pCDNA3, and transfected into CHO cells. After two days, cells were studied by immunofluorescence using monoclonal antibodies against these epitopes and secondary antibodies labelled with Cy2 and Cy3. To determine surface expression of KCNE3 in oocytes, we used HA-tagged KCNE3 and untagged KCNQ1 and the enzymatic detection system as described²¹.

Patch-clamp analysis of colonic crypt cells

We prepared colonic crypts as described²⁸ and performed whole-cell patch clamp measurements. Cells were held at their membrane voltage and a clamp protocol was applied every 15 s as indicated in Fig. 5. Patch pipettes were filled with 95 mM K-gluconate, 30 mM KCl, 1.2 mM NaH₂PO₄, 4.8 mM Na₂HPO₄, 0.73 mM Ca-gluconate, 1 mM MgCl₂, 1 mM EGTA, 5 mM D-glucose and 3 mM ATP. The bath solution contained 145 mM Na-gluconate, 1.6 mM K₂HPO₄, 0.4 mM KH₂PO₄, 8 mM Ca-gluconate, 1 mM MgSO₄ and 5 mM D-glucose. cAMP was raised by applying 5 μM forskolin and 100 μM 8-CPT cAMP to the bath. In some experiments the Ca²⁺-activated K⁺-channel was additionally inhibited by 5 nM charybdotoxin.

Received 19 October; accepted 24 November 1999.

- Wang, Q. *et al.* Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nature Genet.* **12**, 17–23 (1996).
- Neyroud, N. *et al.* A novel mutation in the potassium channel gene KVLQT1 causes the Jervell and Lange-Nielsen cardioauditory syndrome. *Nature Genet.* **15**, 186–189 (1997).
- Bievart, C. *et al.* A potassium channel mutation in neonatal human epilepsy. *Science* **279**, 403–406 (1998).
- Singh, N. A. *et al.* A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. *Nature Genet.* **18**, 25–29 (1998).
- Charlier, C. *et al.* A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. *Nature Genet.* **18**, 53–55 (1998).
- Kubisch, C. *et al.* KCNQ4, a novel potassium channel expressed in sensory outer hair cells, is mutated in dominant deafness. *Cell* **96**, 437–446 (1999).
- Takumi, T., Ohkubo, H. & Nakanishi, S. Cloning of a membrane protein that induces a slow voltage-gated potassium current. *Science* **242**, 1042–1045 (1988).
- Barhanin, J. *et al.* KvLQT1 and IsK (minK) proteins associate to form the I_{Ks} cardiac potassium current. *Nature* **384**, 78–80 (1996).
- Sanguinetti, M. C. *et al.* Coassembly of KvLQT1 and minK (IsK) proteins to form cardiac I_{Ks} potassium channel. *Nature* **384**, 80–83 (1996).
- Chouabe, C. *et al.* Properties of KvLQT1 K⁺ channel mutations in Romano-Ward and Jervell and Lange-Nielsen inherited cardiac arrhythmias. *EMBO J.* **16**, 5472–5479 (1997).
- Wollnik, B. *et al.* Pathophysiological mechanisms of dominant and recessive KvLQT1 K⁺ channel mutations found in inherited cardiac arrhythmias. *Hum. Mol. Genet.* **6**, 1943–1949 (1997).
- Splawski, J., Tristani-Firouzi, M., Lehmann, M. H., Sanguinetti, M. C. & Keating, M. T. Mutations in the hminK gene cause long QT syndrome and suppress I_{Ks} function. *Nature Genet.* **17**, 338–340 (1997).
- Offner, F. F., Dallos, P. & Cheatham, M. A. Positive endocochlear potential: mechanism of production by marginal cells of stria vascularis. *Hear. Res.* **29**, 117–124 (1987).

14. Suessbrich, H. Specific blockade of slowly activating I_{Ks} channels by chromanols—impact on the role of I_{Ks} channels in epithelia. *FEBS Lett.* **396**, 271–275 (1996).
15. Abbott, G. W. *et al.* MiRP1 forms I_{Ks} potassium channels with HERG and is associated with cardiac arrhythmia. *Cell* **97**, 175–187 (1999).
16. Busch, A. E. *et al.* The role of the I_{Ks} protein in the specific pharmacological properties of the I_{Ks} channel complex. *Br. J. Pharmacol.* **122**, 187–189 (1997).
17. Devor, D. C., Singh, A. K., Gerlach, A. C., Frizzell, R. A. & Bridges, R. J. Inhibition of intestinal Cl^- secretion by clonidine: direct effect on basolateral membrane K^+ channels. *Am. J. Physiol.* **273**, C531–C540 (1997).
18. Rufo, P. A. *et al.* The antifungal antibiotic, clotrimazole, inhibits chloride secretion by human intestinal T84 cells via blockade of distinct basolateral K^+ conductances. Demonstration of efficacy in intact rabbit colon and in an *in vivo* mouse model of cholera. *J. Clin. Invest.* **100**, 3111–3120 (1997).
19. Yang, W. P. *et al.* KvLQT1, a voltage-gated potassium channel responsible for human cardiac arrhythmias. *Proc. Natl Acad. Sci. USA* **94**, 4017–4021 (1997).
20. Schroeder, B. C., Kubisch, C., Stein, V. & Jentsch, T. J. Moderate loss of function of cyclic-AMP-modulated KCNQ2/KCNQ3 K^+ channels causes epilepsy. *Nature* **396**, 687–690 (1998).
21. Zerangue, N., Schwappach, B., Jan, Y. N. & Jan, L. Y. A new ER trafficking signal regulates the subunit stoichiometry of plasma membrane K_{ATP} channels. *Neuron* **22**, 537–548 (1999).
22. McDonald, T. V. *et al.* A minK-HERG complex regulates the cardiac potassium current I_{Kr} . *Nature* **388**, 289–292 (1997).
23. Curran, M. E. *et al.* A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell* **80**, 795–803 (1995).
24. Lohrmann, E. *et al.* A new class of inhibitors of cAMP-mediated Cl^- secretion in rabbit colon, acting by the reduction of cAMP-activated K^+ conductance. *Pflügers Arch.* **429**, 517–530 (1995).
25. Devor, D. C., Singh, A. K., Frizzell, R. A. & Bridges, R. J. Modulation of Cl^- secretion by benzimidazolones. I. Direct activation of a Ca^{2+} -dependent K^+ channel. *Am. J. Physiol.* **271**, L775–L784 (1996).
26. Mall, M. *et al.* Cholinergic ion secretion in human colon requires coactivation by cAMP. *Am. J. Physiol.* **275**, G1274–G1281 (1998).
27. MacVinish, L. J., Hickman, M. E., Mufti, D. A., Durrington, H. J. & Cuthbert, A. W. Importance of basolateral K^+ conductance in maintaining Cl^- secretion in murine nasal and colonic epithelia. *J. Physiol. (Lond.)* **510**, 237–247 (1998).
28. Warth, R. *et al.* The cAMP-regulated and 293B-inhibited K^+ conductance of rat colonic crypt base cells. *Pflügers Arch.* **432**, 81–88 (1996).
29. Diener, M., Hug, F., Strabel, D. & Scharrer, E. Cyclic AMP-dependent regulation of K^+ transport in the rat distal colon. *Br. J. Pharmacol.* **118**, 1477–1487 (1996).
30. Trezise, A. E. & Buchwald, M. *In vivo* cell-specific expression of the cystic fibrosis transmembrane conductance regulator. *Nature* **353**, 434–437 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. P. Hardelin and C. Petit for performing KCNE3 *in situ* hybridization of cochlear sections. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie to R.G. and T.J.J.

Correspondence and requests for materials should be addressed to T.J.J. (e-mail: jentsch@plexus.uke.uni-hamburg.de).

Uptake of apoptotic cells drives the growth of a pathogenic trypanosome in macrophages

Célio G. Freire-de-Lima*, Danielle O. Nascimento*, Milena B. P. Soares†, Patricia T. Bozza‡, Hugo C. Castro-Faria-Neto‡, Fernando G. de Mello*, George A. DosReis* & Marcela F. Lopes*

* Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, 21944-970, Rio de Janeiro, RJ, Brazil

† Centro de Pesquisas Gonçalo Moniz, FIOCRUZ, 40295-001, Salvador, BA, Brazil

‡ Instituto Oswaldo Cruz, FIOCRUZ, 21045-900, Rio de Janeiro, RJ, Brazil

After apoptosis, phagocytes prevent inflammation and tissue damage by the uptake and removal of dead cells¹. In addition, apoptotic cells evoke an anti-inflammatory response through macrophages^{2,3}. We have previously shown that there is intense lymphocyte apoptosis in an experimental model of Chagas' disease⁴, a debilitating cardiac illness caused by the protozoan *Trypanosoma cruzi*. Here we show that the interaction of apoptotic

cells, but not necrotic T lymphocytes with macrophages infected with *T. cruzi* fuels parasite growth in a manner dependent on prostaglandins, transforming growth factor- β (TGF- β) and polyamine biosynthesis. We show that the vitronectin receptor is critical, in both apoptotic-cell cytoadherence and the induction of prostaglandin E_2 /TGF- β release and ornithine decarboxylase activity in macrophages. A single injection of apoptotic cells in infected mice increases parasitaemia, whereas treatment with cyclooxygenase inhibitors almost completely ablates it *in vivo*. These results suggest that continual lymphocyte apoptosis and phagocytosis of apoptotic cells by macrophages have a role in parasite persistence in the host, and that cyclooxygenase inhibitors have potential therapeutic application in the control of parasite replication and spread in Chagas' disease.

We have already shown that the onset of activation-induced cell death in $CD4^+$ T cells exacerbates parasite replication in co-cultured macrophages infected with *T. cruzi*⁵. To investigate whether the clearance of apoptotic cells predisposes macrophages to *T. cruzi* infection, murine resident peritoneal macrophages were exposed to apoptotic, necrotic or viable splenic T cells first, and then washed and infected. Apoptotic, but not necrotic or living T cells increased *T. cruzi* growth in macrophage cultures (Fig. 1a). Similar results were obtained when apoptotic or necrotic cells were added after *T. cruzi* infection (data not shown). Nevertheless, treatment of lymphocytes with the caspase-inhibitor zVAD-fmk peptide before apoptosis induction, rescued T cells from death (data not shown) and prevented the increase in parasite replication (Fig. 1b) in a dose-dependent manner. In another model, peritoneal macrophages from mice infected with *T. cruzi* were incubated with apoptotic or necrotic cells. Apoptotic, but not necrotic cells also exacerbated endogenous *T. cruzi* growth in these *in vivo* infected macrophages (Fig. 1c). In agreement with *in vitro* results, a single *in vivo* injection of apoptotic, but not necrotic cells in *T. cruzi*-infected mice resulted in a sudden rise in parasitaemia (Fig. 1d).

Previous studies pointed to a role for an integrin in the recognition and ingestion of apoptotic cells by macrophages^{6,7}. We observed that RGDS, but not RGE peptide blocked apoptotic cell binding and reduced *T. cruzi* growth within macrophages (data not shown).

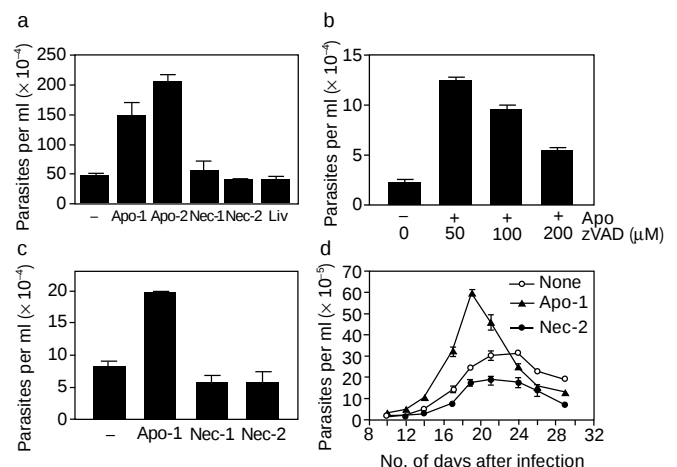


Figure 1 Apoptotic cells exacerbate parasite growth in *T. cruzi* infection. **a**, Peritoneal resident macrophages were either untreated (-) or exposed to Apo-1, Apo-2, Nec-1, Nec-2 or to living cells (Liv). After 5 days, cells were removed and macrophages were infected with *T. cruzi*; parasites were counted 10 days later. **b**, T cells (Apo) were treated with the indicated doses of zVAD-fmk (zVAD) before apoptosis induction (by heating) and added to macrophages infected *in vitro*. Parasites were counted 7 days later. Maximal cell death (100%) and increased parasite replication still occurred with zVAD-fmk at 50 μM. **c**, Macrophages from infected mice were exposed to Apo-1, Nec-1 or Nec-2 throughout culture, and trypomastigotes counted 10 days later. **d**, Mice were injected 7 days after infection with Apo-1 or Nec-2, and parasitaemia was monitored.

The vitronectin receptor (VnR) integrin ($\alpha_v\beta_3$, CD51/CD61) is involved in phagocytosis of apoptotic cells by macrophages^{6,7}. Both VnR and CD36 take part in the macrophage receptor that bridges thrombospondin exposed on apoptotic cells⁷. Flow cytometry analysis revealed that both α_v - and β_3 -integrin chains were absent from the surface of control resident macrophages, but their expression was upregulated in the course of *T. cruzi* infection (Fig. 2a, b). In addition, $\alpha_v\beta_3$ expression was induced in control macrophages after a 24-h culture (not shown). We therefore investigated the role of VnR in the binding of apoptotic cells by *in vivo* infected macrophages. After a 3-h incubation, nearly all macrophages were rosetted by apoptotic cells (mean $93.5 \pm 2.5\%$ positive macro-

phages). Both intact monoclonal antibodies against α_v - or β_3 -VnR chains and anti- α_v fragment of antigen binding (Fab) inhibited apoptotic cell binding by 40–50% (Fig. 2c).

We then investigated VnR involvement in parasite replication. Engagement of VnR by anti- α_v or anti- β_3 monoclonal antibodies, either soluble (Fig. 2d) or immobilized on plates (Fig. 3c), markedly increased *T. cruzi* growth in macrophages in the absence of apoptotic cells. Soluble anti- α_v Fab fragments failed to enhance parasite replication, but promoted *T. cruzi* growth upon crosslinkage by a secondary goat anti-hamster IgG (Fig. 2e). Most notably, soluble anti- α_v Fab completely blocked the apoptotic cell effect on parasite replication in macrophages obtained from infected mice (Fig. 2f). A control hamster anti-CD69 Fab had no effect (Fab-1 in Fig. 2f), even though it completely inhibited CD69 staining in the macrophages used in these experiments (data not shown). Overall, these results show that blockade of VnR was sufficient to ablate parasite replication, whereas VnR crosslinkage mimicked the effects of apoptotic cells on *T. cruzi* growth in macrophages.

Interaction with apoptotic cells leads macrophages to secrete TGF- β , which in turn suppresses their pro-inflammatory cytokine response³. In addition, TGF- β renders both phagocytic and non-phagocytic cells permissive to *T. cruzi* infection^{8,9} and antagonizes interferon (IFN)- γ -induced nitric oxide (NO) production and trypanocidal activity of macrophages¹⁰. We investigated the role of TGF- β in *T. cruzi* infection in macrophages treated with apoptotic cells. Both uninfected and infected macrophages produced similar levels of TGF- β in response to apoptotic, but not necrotic, cells (Fig. 3a). We also detected marked secretion of TGF- β triggered by anti- α_v monoclonal antibody (Fig. 3b) in the absence of *T. cruzi* infection and apoptotic cells. Anti- α_v Fab promoted TGF- β production only in the presence of a secondary goat anti-hamster IgG (Fig. 3b), but completely ablated TGF- β secretion in response to apoptotic cells (data not shown). We treated infected macrophages with anti-TGF- β 1 neutralizing antibody. Neutralization of TGF- β inhibited parasite replication induced by either apoptotic cells or immobilized anti- β_3 monoclonal antibody (Fig. 3c), whereas a control IgY antibody had no effect. Although the involvement of other co-receptors or co-factors² cannot be ruled out, the present data indicate that engagement of VnR by apoptotic cells may result in a TGF- β -dependent increase in parasite replication in macrophages. We also investigated whether apoptotic cells and TGF- β interfere with NO-dependent trypanocidal activity of macrophages. Apoptotic cells blocked NO production by IFN- γ /LPS-activated macrophages (Fig. 3d, left) and led to vigorous *T. cruzi* replication (Fig. 3d, right). Inhibition of NO production by apoptotic cells could be reversed by anti-TGF- β antibodies or anti- α_v Fab (Fig. 3e). Therefore, the uptake of apoptotic cells renders macrophages refractory to inflammatory cytokines, allowing parasite survival and growth even in the face of immune response.

Transforming growth factor- β shifts arginine metabolism in macrophages, decreasing NO and inducing ornithine production (by arginase) and subsequent polyamine biosynthesis¹¹. Ornithine decarboxylase (ODC) catalyses putrescine synthesis from ornithine, and is considered as the limiting step in the polyamine biosynthetic pathway, leading to spermidine and spermine production¹². Although ODC activity has not been identified in *T. cruzi*, intra- and extracellular forms of *T. cruzi* synthesize polyamines either through the alternative agmatine pathway¹³ or by using exogenously added putrescine¹⁴. To investigate the possibility that macrophage putrescine synthesis is involved in the apoptotic cell effects that we observed, putrescine content and ODC activity were measured in macrophages treated with apoptotic cells. Macrophages treated with apoptotic cells markedly accumulated putrescine (Fig. 3f, left), and exogenous addition of putrescine alone increased parasite replication in macrophages (Fig. 3f, right). ODC activity was markedly upregulated in macrophages exposed to apoptotic cells or to immobilized anti- α_v monoclonal antibody, with delayed kinetics

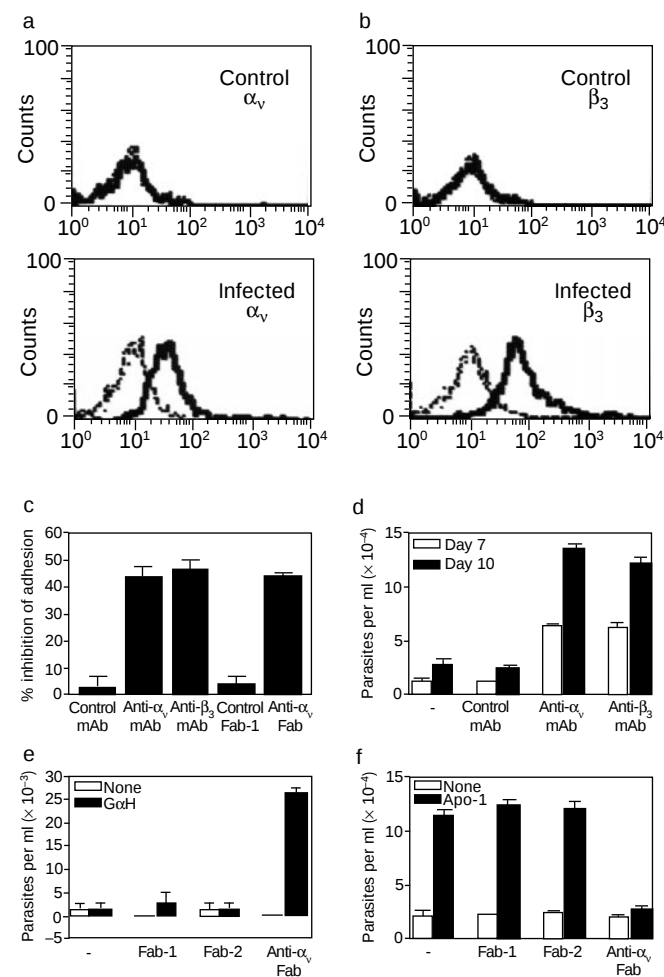


Figure 2 Effects of apoptotic cells on macrophages infected with *T. cruzi* are mediated by the vitronectin receptor. α_v (a) and β_3 (b) VnR chains are expressed by macrophages from infected (lower panels), but not uninfected (upper panels), mice. Mac-1⁺ cells were gated and analysed for α_v (a) or β_3 (b) expression. Bold lines, anti- α_v (a) or anti- β_3 (b) staining; dotted lines, cells stained after treatment with unlabelled anti- α_v (a) or anti- β_3 (b) monoclonal antibodies; dashed lines, unstained cells. c, Anti-VnR monoclonal antibodies and Fab inhibit apoptotic cell binding to macrophages. Macrophages infected *in vivo* were treated with anti- α_v Fab or control anti-CD69 Fab-1, anti- α_v , anti- β_3 or control hamster IgG monoclonal antibodies (mAb) (20 min), and then exposed to Apo-1. Adhesion was determined after 3 h. d, e, Anti- α_v or anti- β_3 , but not control monoclonal antibody drives *T. cruzi* replication within macrophages. Macrophages infected *in vivo* were treated with monoclonal antibodies (d) or were treated first with anti- α_v Fab, Fab-1 or control hamster IgG Fab-2 (50 μ g ml⁻¹) for 20 min, washed and treated with goat anti-hamster IgG antibody (G α H) at 10 μ g ml⁻¹ (e). Parasites were counted as indicated (d) or after 10 days (e). f, Anti- α_v Fab inhibits the apoptotic cell effects on *T. cruzi* replication. Macrophages infected *in vivo* were treated first with Fab fragments (20 min) and then with Apo-1. Parasites were counted 10 days later.

(Fig. 3g). Both anti- α_v Fab and anti-TGF- β antibodies suppressed apoptotic cell effects on ODC activity (Fig. 3h). A competitive ODC inhibitor, α -methylornithine¹⁵, also blocked ODC activity induced by apoptotic cells (Fig. 3g). Addition of this inhibitor to macrophages treated with apoptotic cells or anti- α_v resulted in a dose-dependent decrease in parasite replication (Fig. 3i), but it had no inhibitory effect on basal *T. cruzi* growth in macrophages left without stimuli (Fig. 3i), and did not affect the uptake of apoptotic cells or TGF- β production (data not shown). These results indicate that TGF- β , produced upon engagement of VnR by apoptotic cells, may induce ODC activity in macrophages and promote polyamine-dependent parasite growth.

Searching therapeutic targets to prevent apoptotic cell effects, a role for prostaglandins on *T. cruzi* growth was investigated. Prostaglandin E_2 and PAF have been reported to induce TGF- β production by human macrophages exposed to apoptotic neutrophils³; prostaglandin E_2 increases arginase¹⁶ and ODC¹⁷ activities, and promoted *T. cruzi* growth (data not shown) in macrophages. We

therefore evaluated PGE₂ production by macrophages exposed to either apoptotic cells or immobilized anti- α_v monoclonal antibodies. Apoptotic cells induced PGE₂ production by macrophages, and this effect was prevented by anti- α_v Fab but not by control anti-CD69 Fab-1 (Fig. 4a). Vitronectin receptor ligation by immobilized anti- α_v monoclonal antibodies also increased PGE₂ levels, whereas anti-CD69 or anti-Mac-1 (α_M/β_2 integrin) control monoclonal antibodies had no effect (Fig. 4a). We tested three non-steroidal anti-inflammatory drugs (NSAIDs): aspirin, an inhibitor of both constitutive and inducible cyclooxygenases (COX); indomethacin, a preferential antagonist of constitutive COX¹⁸; and NS-398, an inducible COX inhibitor¹⁹. Indomethacin almost completely blocked anti-VnR induced PGE₂ (Fig. 4a) and TGF- β (data not shown) production. Indomethacin, aspirin and NS-398 significantly suppressed apoptotic cell effects on PGE₂ (Fig. 4b) and TGF- β (data not shown) secretion, on ODC activity *in vitro* (Fig. 4c) and *in vivo* (Fig. 4d), and on *T. cruzi* growth, without affecting basal parasite replication (Fig. 4e). Furthermore, COX

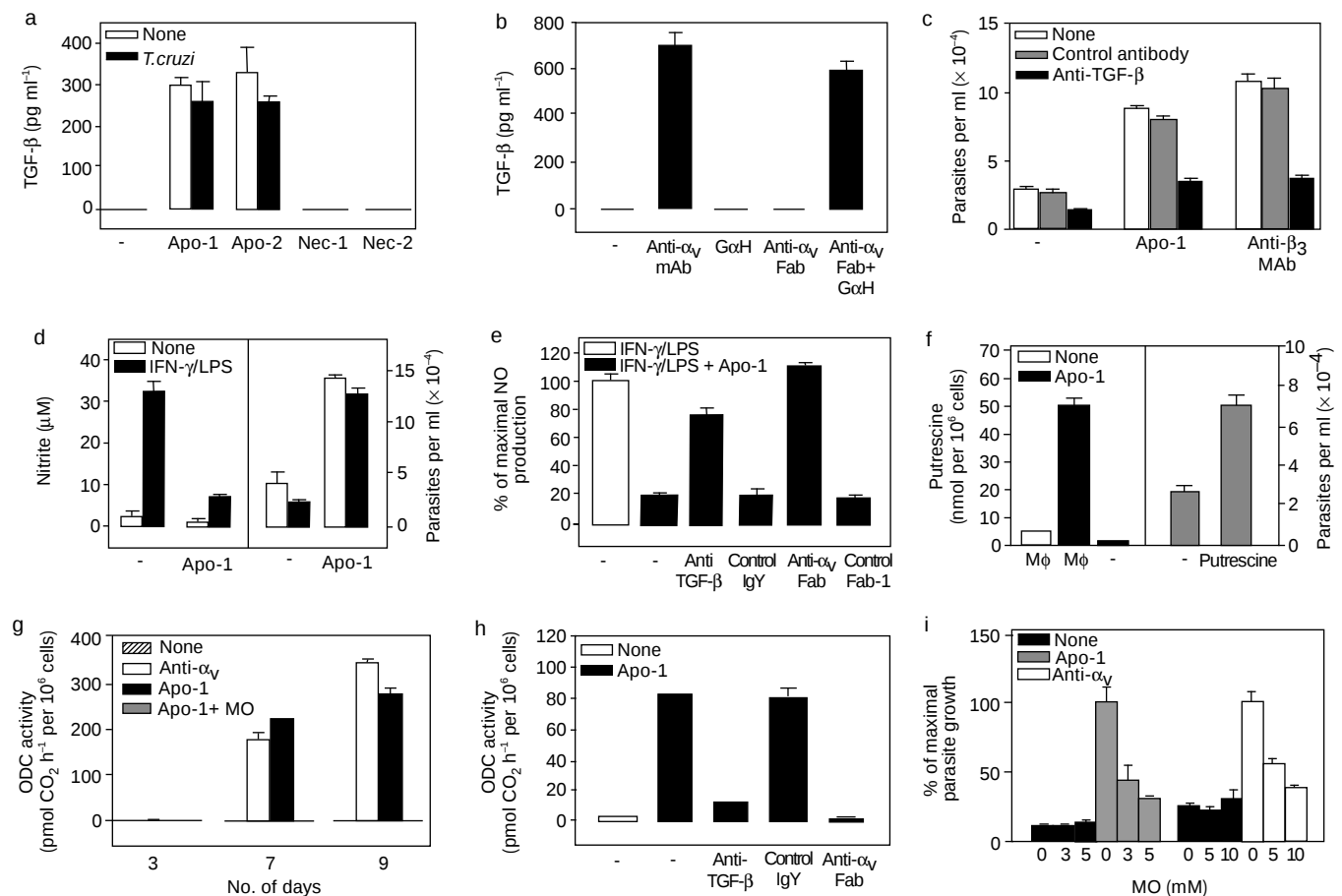


Figure 3 Effects of apoptotic cells on macrophages infected with *T. cruzi* depend on TGF- β . **a**, Apoptotic (Apo-1, -2), but not necrotic (Nec-1, -2), cells induce TGF- β production by uninfected (open bars) or *in vitro* infected (closed bars) macrophages. TGF- β 1 was evaluated in supernatants from 48-h cultures. **b**, VnR engagement induces TGF- β production by macrophages. Uninfected macrophages were exposed to anti- α_v monoclonal antibody (mAb) or Fab, followed or not by goat anti-hamster IgG (G α H), and TGF- β 1 was evaluated in supernatants from 24-h cultures. **c**, Neutralization of TGF- β ablates the effect of apoptotic cells or anti-VnR mAb on macrophages infected with *T. cruzi*. *In vivo* infected macrophages were exposed to Apo-1 or to immobilized anti- β_3 MAb in the presence of either anti-hTGF- β 1 (6 μ g ml⁻¹) or control IgY antibodies. Trypomastigotes were counted after 10 days. **d**, **e**, Apoptotic cells suppress NO production by LPS/IFN- γ -activated macrophages. *In vivo* infected (**d**) or uninfected (**e**) macrophages were exposed to LPS plus IFN- γ and treated with Apo-1, anti-TGF- β or control IgY antibody, anti- α_v Fab or anti-CD69 Fab-1. Supernatants were assayed for

nitrite content after 48 h (**d**, left, **e**) and for parasites after 10 days (**d**, right). In **e**, results were expressed as a percentage of the maximal nitrite content in macrophages treated with LPS plus IFN- γ . **f**, Apoptotic cells (Apo-1) induce putrescine accumulation in uninfected macrophages (M ϕ) and exogenous putrescine (1 mM) increases parasite growth in macrophages infected *in vitro*. Putrescine (left) and parasite accumulation (right) were measured after 7 days. **g**, Kinetics of ODC induction. Macrophages were exposed to immobilized anti- α_v or to Apo-1. ODC inhibitor MO was added at 10 mM. ODC activity was measured at the indicated days. **h**, Apoptotic cell effects on ODC activity depend on TGF- β . Macrophages were exposed to Apo-1 in the presence of anti- α_v Fab, anti-TGF- β or control IgY antibodies. ODC activity was evaluated after 7 days. **i**, ODC activity is required for *T. cruzi* growth. *In vitro* infected macrophages were exposed to Apo-1 or to anti- α_v mAb, and MO was added at the indicated dosages. Trypomastigotes were counted after 8 (anti- α_v) or 10 (Apo-1) days in culture and are expressed as a percentage of the maximal parasite growth in the absence of MO.

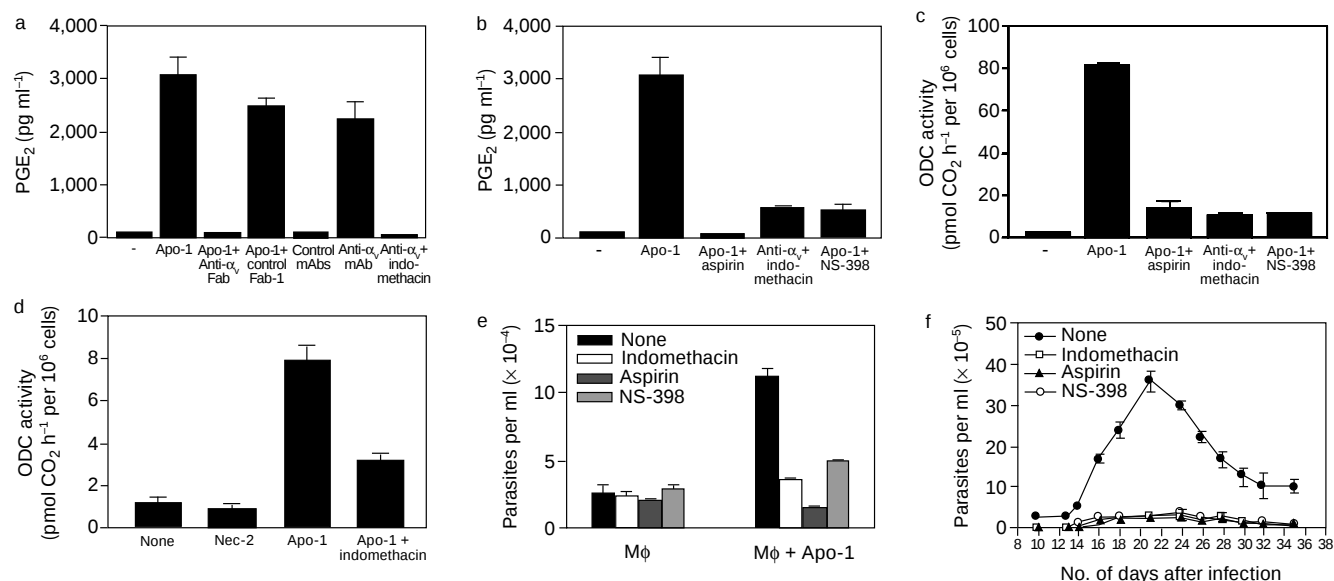


Figure 4 *In vitro* and *in vivo* effects of COX inhibitors on *T. cruzi* infection. **a**, Apoptotic cells or VnR ligation induce PGE₂ release by macrophages. Macrophages were treated with Apo-1, anti-α_v Fab or anti-CD69 Fab-1, immobilized anti-α_v, anti-CD69 or anti-Mac-1 mAbs, or anti-α_v plus indomethacin. PGE₂ was measured as described. For simplicity, results with anti-CD69 and anti-Mac-1 were combined in the figure (control mAbs). **b**, **c**, COX inhibitors block PGE₂ release and ODC activity *in vitro*. Macrophages were exposed to Apo-1 and treated with aspirin, indomethacin or NS-398. PGE₂ levels were measured after 24 h (**b**), and ODC activity was assayed after 7 days (**c**). **d**, Apoptotic cells

induce ODC activity *in vivo*. Mice (*n* = 3) were injected i.p. with Nec-2, Apo-1 or Apo-1 plus indomethacin. ODC activity was measured 10 days later in adherent PEC. **e**, COX inhibitors antagonize apoptotic cell effects on macrophages infected with *T. cruzi*. *In vitro* infected macrophages (Mφ) were exposed to Apo-1 and treated with COX inhibitors as in **b**. Trypomastigotes were counted after 7 days. **f**, COX inhibitors suppress parasitaemia in *T. cruzi* infected mice. Mice (*n* = 4) were infected with *T. cruzi* and treated 7, 8 and 9 days after infection with aspirin, indomethacin or NS-398. Parasitaemia was monitored throughout acute infection.

inhibitors almost completely abolished parasitaemia in infected mice, after a brief 3-day treatment, one week after infection (Fig. 4f). These data indicate that the factors driving *T. cruzi* replication *in vivo* may be the same mediators as are involved in the parasite growth-promoting activity of apoptotic cells *in vitro*. However, these results should be taken with caution, as NSAIDs can have differential effects on macrophages depending on dosage²⁰. More experiments are necessary to understand the discrepant effects of high-dose therapy with COX inhibitors seen in other reported studies on *T. cruzi* infection²¹.

Our results show that TGF-β engagement by apoptotic cells triggers PGE₂ and TGF-β release by macrophages. The suppressive effect of PGE₂/TGF-β on pro-inflammatory cytokine expression³ and NO production would create the appropriate environment for optimal *T. cruzi* growth within macrophages. Furthermore, the uptake of apoptotic cells reorients macrophage metabolism towards putrescine production, helping parasite replication. We have shown that widely used drugs, such as aspirin and indomethacin, interfere with this pathway and are able to control parasitaemia in susceptible mice, making them potentially useful for therapy in the acute phase of Chagas' disease. □

Methods

Animals and infection

We infected BALB/c male mice (6 weeks of age) intraperitoneally (i.p.) with 10⁴ metacyclic trypomastigote forms of *T. cruzi* clone Dm 28c (ref. 22).

Macrophages

Peritoneal exudate cells (PEC) removed from the peritoneal cavity of uninfected or acutely infected mice (3–4 weeks after infection) were cultured in complete medium (DMEM, 2 mM L-glutamine, 1 mM sodium pyruvate, 10 μg ml⁻¹ gentamicin, MEM non-essential amino acids, 10 mM Hepes and 50 μM 2-ME) with 1% Nutridoma-SP (Boehringer Mannheim) or 5% FCS (experiments in Fig. 1a, c) at 3 × 10⁵ cells ml⁻¹ on 24-well plates (Corning). Macrophages from normal mice were infected with 1.5 × 10⁶ *T. cruzi* metacyclic forms per well⁵. After 24 h, non internalized parasites were removed and macrophages cultured in complete medium (1 ml) at 37 °C, with 7% CO₂ for up to 10 days. Extracellular motile trypomastigotes were counted in culture supernatants⁵.

Lymphocytes

Nylon-wool-filtered normal splenocytes were suspended in complete medium and heated (56 °C for 30 min) (Apo-1) or irradiated with 30 Gy (Apo-2) to obtain apoptotic cells as described²³. Cells were also fixed with 1% paraformaldehyde for 20 min at room temperature (followed by extensive washing) (Nec-1) or frozen-thawed (Nec-2) to obtain necrotic cells²³. Dead or viable cells (10⁶ per well) were then added to macrophages.

Antibodies and reagents

Anti-β₃ integrin (CD61), PE-labelled anti-β₃ and anti-α_v (CD51), control hamster IgG, unlabelled and FITC- and PE-labelled anti-Mac-1, FITC-labelled anti-CD69, and anti-CD16/CD32 monoclonal antibodies were purchased from Pharmingen (San Diego). Purified chicken anti-hTGF-β1 IgY was from R&D Systems (Minneapolis) and control (chicken anti-canine IgG) IgY antibody (a gift from V. Laurentino) was prepared as described²⁴. Anti-α_v mAb H9.2B8 (ref. 25), and anti-CD69 mAb H1.2F3 (ref. 26) were donated by E. Shevach. Fab fragments of control hamster IgG, anti-CD69 and anti-α_v monoclonal antibodies were produced with a commercial Fab preparation kit (Pierce). zVAD-fmk peptide was a gift from M. Lenardo. Other reagents were goat anti-hamster IgG antibody (Cappel-Organon Teknica Corporation), mrIFN-γ (Pharmingen), LPS (*Escherichia coli* O111:B4; Difco), α-methyl-ornithine, MO (Marion Merrel Dow), Putrescine (Sigma), aspirin and indomethacin (Sigma), and NS-398 (Biomol).

Adhesion assay

Macrophages were cultured with Apo-1 and anti-α_v, anti-β₃ or control hamster IgG mAbs (10 μg ml⁻¹), anti-α_v Fab or control anti-CD69 Fab-1 (50 μg ml⁻¹) for 3 h, washed, fixed with 1% paraformaldehyde, and counted under phase-contrast microscopy. Adhesion was calculated as the percentage of rosetted macrophages out of a total of 100 macrophages counted per well.

TGF-β production

Macrophages were treated with apoptotic or necrotic cells, or with anti-α_v monoclonal antibody or Fab plus goat anti-hamster IgG (10 μg ml⁻¹). The content of TGF-β1 in supernatants was evaluated by sandwich ELISA as described²⁷.

NO production

Macrophages were cultured with IFN-γ (40 U ml⁻¹) plus LPS (10 ng ml⁻¹), and exposed or not to Apo-1, anti-TGF-β or control IgY antibody (6 μg ml⁻¹), anti-α_v Fab or control anti-CD69 Fab-1. Supernatants were collected 48 h later and mixed with an equal volume of Griess reagent to determine nitrite content, as described²⁸.

PGE₂ release

Macrophages were exposed to immobilized anti-α_v or control monoclonal antibodies (10 μg ml⁻¹), Apo-1, Fab fragments, aspirin (10 μg ml⁻¹), indomethacin (1 μg ml⁻¹) or

NS-398 (1 μ M). Culture supernatants (24 h) were collected and assayed for PGE₂ using a competitive ELISA kit (Cayman Chemicals).

ODC activity and putrescine content

Macrophages were cultured (10⁶ per ml) with Apo-1 (3 $\times$ 10⁶ per well), anti- α , monoclonal antibody, MO or other reagents. Ornithine decarboxylase activity and putrescine content were evaluated as described²⁹.

Flow cytometry

Peritoneal exudate cells (PEC) were treated with Fc block (anti-CD16/CD32) and stained with FITC-labelled anti-Mac-1 plus PE-labelled anti- α , or anti- β 3 monoclonal antibodies. Unlabelled anti- α , or anti- β 3 monoclonal antibodies were used to block staining specifically. Uninfected macrophages were also cultured for 24 h, detached and stained. Antibodies were used at 1 μ g per 10⁶ cells. 10⁴ cells were acquired, and Mac-1-positive cells were gated and analysed for either α , or β 3 expression on a B-D Xcalibur flow cytometer.

In vivo experiments

Infected mice (n = 4) were injected i.p. with 10⁷ Apo-1 or Nec-2, 7 days after infection, or left without treatment. Parasitaemia was determined on blood samples from the tail. In other experiments, infected mice (n = 4) were injected i.p. at 7, 8 and 9 days after infection, with aspirin (50 mg kg⁻¹), indomethacin (1 mg kg⁻¹) or NS-398 (5 mg kg⁻¹), or left untreated, and parasitaemia was followed. For *ex-vivo* determination of ODC activity, mice (n = 3) were injected i.p. with 10⁷ Apo-1 or Nec-2, or left without treatment. Mice injected with Apo-1 were untreated or treated with indomethacin (1 mg kg⁻¹) in the same day, and 4 days later. ODC activity was measured in adherent PEC (10⁶ per well) 10 days later.

Presentation of results and statistics

Each experiment presented is representative of at least three independent experiments. Data are expressed as mean $\pm$ s.e. of duplicate determinations. For *in vivo* experiments, results are expressed as mean $\pm$ s.e. of individual animals. Significance was evaluated by Student's unpaired *t*-test, and all positive results mentioned were significant (P < 0.05 or <0.01) compared with controls.

Received 13 September; accepted 5 November 1999.

- Savill, J. Apoptosis: Phagocytic docking without shocking. *Nature* **392**, 442–443 (1998).
- Voll, R. E., Herrmann, M., Roth, E. A., Stach, C. & Kalden, J. R. Immunosuppressive effects of apoptotic cells. *Nature* **390**, 350–351 (1997).
- Fadok, V. A. *et al.* Macrophages that have ingested apoptotic cells *in vitro* inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF- β , PGE₂ and PAF. *J. Clin. Invest.* **101**, 890–898 (1998).
- Lopes, M. F., Veiga, V. F., Santos, A. R., Fonseca, M. E. F. & DosReis, G. A. Activation-induced CD4⁺ T cell death by apoptosis in experimental Chagas disease. *J. Immunol.* **154**, 744–752 (1995).
- Nunes, M. P., Andrade, R. M., Lopes, M. F. & DosReis, G. A. Activation-induced T cell death exacerbates *Trypanosoma cruzi* replication in macrophages cocultured with CD4⁺ T lymphocytes from infected hosts. *J. Immunol.* **160**, 1313–1319 (1998).
- Savill, J., Dransfield, I., Hogg, N. & Haslett, C. Vitronectin receptor-mediated phagocytosis of cells undergoing apoptosis. *Nature* **343**, 170–173 (1990).
- Savill, J., Hogg, N., Ren, Y. & Haslett, C. Thrombospondin cooperates with CD36 and the vitronectin receptor in macrophage recognition of neutrophils undergoing apoptosis. *J. Clin. Invest.* **90**, 1513–1522 (1992).
- Silva, J. S., Twardzik, D. R. & Reed, S. G. Regulation of *Trypanosoma cruzi* infection *in vitro* and *in vivo* by transforming growth factor β (TGF- β). *J. Exp. Med.* **174**, 539–545 (1991).
- Ming, M., Ewen, M. E. & Pereira, M. E. A. Trypanosome invasion of mammalian cells requires activation of the TGF- β signaling pathway. *Cell* **82**, 287–296 (1995).
- Gazzinelli, R. T., Oswald, I. P., Hieny, S., James, S. L. & Sher, A. The microbicidal activity of interferon- γ -treated macrophages against *Trypanosoma cruzi* involves an L-arginine-dependent, nitrogen oxide-mediated mechanism inhibitable by interleukin-10 and transforming growth factor- β . *Eur. J. Immunol.* **22**, 2501–2506 (1992).
- Boutard, V. *et al.* Transforming growth factor- β stimulates arginase activity in macrophages: implications for the regulation of macrophage cytotoxicity. *J. Immunol.* **155**, 2077–2084 (1995).
- Pegg, A. E. & McCann, P. P. Polyamine metabolism and function. *Am. J. Physiol.* **243**, C212–221 (1982).
- Kierszenbaum, F., Wirth, J. J., McCann, P. P. & Sjoerdsma, A. Arginine decarboxylase inhibitors reduce the capacity of *Trypanosoma cruzi* to infect and multiply in mammalian host cells. *Proc. Natl Acad. Sci. USA* **84**, 4278–4282 (1987).
- Hunter, K. J., Le Quesne, S. A. & Fairlamb, A. H. Identification and biosynthesis of N¹,N⁹-bis(glutathionyl)aminopropylcadaverine (homotrypanothione) in *Trypanosoma cruzi*. *Eur. J. Biochem.* **226**, 1019–1027 (1994).
- Mamont, P. S. *et al.* α -Methyl ornithine, a potent competitive inhibitor of ornithine decarboxylase, blocks proliferation of rat hepatoma cells in culture. *Proc. Natl Acad. Sci. USA* **73**, 1626–1630 (1976).
- Corraliza, I. M., Modolelli, M., Ferber, E. & Soler, G. Involvement of protein kinase A in the induction of arginase in murine bone marrow-derived macrophages. *Biochim. Biophys. Acta* **1334**, 123–128 (1997).
- Prosser, F. H. & Wahl, L. M. Involvement of the ornithine decarboxylase pathway in macrophage collagenase production. *Arch. Biochem. Biophys.* **260**, 218–225 (1998).
- Meade, E. A., Smith, W. L. & DeWitt, D. L. Differential inhibition of prostaglandin endoperoxide synthase (cyclooxygenase) isozymes by aspirin and other non-steroidal anti-inflammatory drugs. *J. Biol. Chem.* **268**, 6610–6614 (1993).
- Futaki, N. *et al.* Selective inhibition of NS-398 on prostanoid production in inflamed tissue in rat carrageenan-air-pouch inflammation. *J. Pharm. Pharmacol.* **45**, 753–755 (1993).

- Jiang, C., Ting, A. T. & Seed, B. PPAR- γ agonists inhibit production of monocyte inflammatory cytokines. *Nature* **391**, 82–85 (1998).
- Celentano, A. M. *et al.* PGE₂ involvement in experimental infection with *Trypanosoma cruzi* subpopulations. *Prostaglandins* **49**, 141–153 (1995).
- Contreras, V. T., Salles, J. M., Thomas, N., Morel, C. M. & Goldenberg, S. *In vitro* differentiation of *Trypanosoma cruzi* under chemically defined conditions. *Mol. Biochem. Parasitol.* **16**, 315–327 (1985).
- Griffith, T. S., Yu, X., Herndon, J. M., Green, D. R. & Ferguson, T. A. CD95-induced apoptosis of lymphocytes in an immune privileged site induces immunological tolerance. *Immunity* **5**, 7–16 (1996).
- Sturmer, A. M., Driscoll, D. P. & Jackson-Matthews, D. E. A quantitative immunoassay using chicken antibodies for detection of native and recombinant α -amidating enzyme. *J. Immunol. Methods* **146**, 105–110 (1992).
- Maxfield, S. R. *et al.* Murine T cells express a cell surface receptor for multiple extracellular matrix proteins. Identification and characterization with monoclonal antibodies. *J. Exp. Med.* **169**, 2173–2190 (1989).
- Yokoyama, W. M. *et al.* Characterization of a cell surface-expressed disulfide-linked dimer involved in murine T cell activation. *J. Immunol.* **141**, 369–376 (1988).
- Soares, M. B. P., David, J. R. & Titus, R. G. An *in vitro* model for infection with *Leishmania major* that mimics the immune response in mice. *Infect. Immun.* **65**, 2837–2845 (1997).
- Green, L. C. *et al.* Analysis of nitrate, nitrite and [¹⁵N] nitrate in biological fluids. *Anal. Biochem.* **126**, 131–138 (1982).
- De Mello, F. G., Bachrach, U. & Nirenberg, M. Ornithine and glutamic acid decarboxylase activities in the developing chick retina. *J. Neurochem.* **27**, 847–851 (1976).

Acknowledgement

We thank E. M. Shevach, M. Lenardo and V. Laurentino for the gifts of peptides and antibodies; and M. A. Vannier dos Santos for reading the manuscript and for helpful discussions. This work was supported by CNPq, FAPERJ, FUIB-UFRJ and PRONEX. C.E.L. is a doctoral fellow of Institute of Microbiology (UFRJ).

Correspondence and request for materials should be addressed to M.F.L. (e-mail: mlopes@niaid.nih.gov) or to G.A.D.R. (e-mail: gdosreis@biof.ufrj.br).

The major protein import receptor of plastids is essential for chloroplast biogenesis

Jörg Bauer*†, Kunhua Chen†‡, Andreas Hiltbunner*†, Ernst Wehrli§, Monika Eugster*, Danny Schnell‡ & Felix Kessler*

* Institute of Plant Sciences, Swiss Federal Institute of Technology, Universitätstrasse 2, CH-8092 Zürich, Switzerland

‡ Department of Biological Sciences, Rutgers, The State University of New Jersey, 101 Warren Street, Newark, New Jersey 07102, USA

§ Laboratory for Electronmicroscopy I, Institute of Biochemistry, Swiss Federal Institute of Technology, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland

† These authors contributed equally to this work

Light triggers the developmental programme in plants that leads to the production of photosynthetically active chloroplasts from non-photosynthetic proplastids¹. During this chloroplast biogenesis, the photosynthetic apparatus is rapidly assembled, mostly from nuclear-encoded imported proteins^{2–4}, which are synthesized in the cytosol as precursors with cleavable amino-terminal targeting sequences called transit sequences. Protein translocon complexes at the outer (Toc complex)^{5–7} and inner (Tic complex)^{6,8,9} envelope membranes recognize these transit sequences, leading to the precursors being imported. The Toc complex in the pea consists of three major components, Toc75, Toc34 and Toc159 (formerly termed Toc86)^{6,7,10,11}. Toc159, which is an integral membrane GTPase¹², functions as a transit-sequence receptor^{5–7,13}. Here we show that *Arabidopsis thaliana* Toc159 (atToc159) is essential for the biogenesis of chloroplasts. In an *Arabidopsis* mutant (*ppi2*) that lacks atToc159, photosynthetic proteins that are normally abundant are transcriptionally repressed, and are found in much smaller amounts in the plastids, although *ppi2* does not affect either the expression or the import

of less abundant non-photosynthetic plastid proteins. These findings indicate that atToc159 is required for the quantitative import of photosynthetic proteins. Two proteins that are related to atToc159 (atToc120 and atToc132) probably help to maintain basal protein import in *ppi2*, and so constitute components of alternative, atToc159-independent import pathways.

The identification and proposed functions of the Toc components are based on the analysis of *in vitro* import assays using isolated pea chloroplasts and recombinant precursors⁵⁻⁷, although there is no evidence for the essential roles of the Toc proteins in plastid protein import *in vivo*. The *Arabidopsis thaliana* mutant *ppi1* (for plastid protein import) has a disruption in a gene encoding one of the two homologues of Toc34 (ref. 14), and causes a non-lethal defect in chloroplast development. We investigated Toc159 because its receptor activity indicates that it might be an important component of the import apparatus. To study the function of Toc159 *in vivo*, we used a reverse genetic approach¹⁵ to identify

mutants in *Arabidopsis* *TOC159*. Searches of *Arabidopsis* genomic sequence databases revealed the existence of a family of three putative import receptor proteins related to pea Toc159 (psToc159) (Fig. 1a). These three proteins, which we designate atToc159, atToc132 and atToc120 on the basis of their deduced relative molecular masses, have a tripartite structure (Fig. 1a). The GTP-binding (Fig. 1a, black and white print) and carboxy-terminal membrane anchor domains (Fig. 1a, green) are highly conserved among the three proteins (~65% identity), whereas the N-terminal acidic domains (Fig. 1a, red) vary considerably in sequence (~20% identity) and length. AtToc159 and psToc159 are most closely related (48% overall identity; 74% identity in the GTP-binding and C-terminal domains), indicating that they are functional orthologues^{10,16}. The messenger RNAs for all three proteins are present in *Arabidopsis* seedlings, with atToc159 mRNA being five- to tenfold more abundant than those of atToc132 and atToc120 in both etiolated and green seedlings (Fig. 1b). Strikingly, the messages for

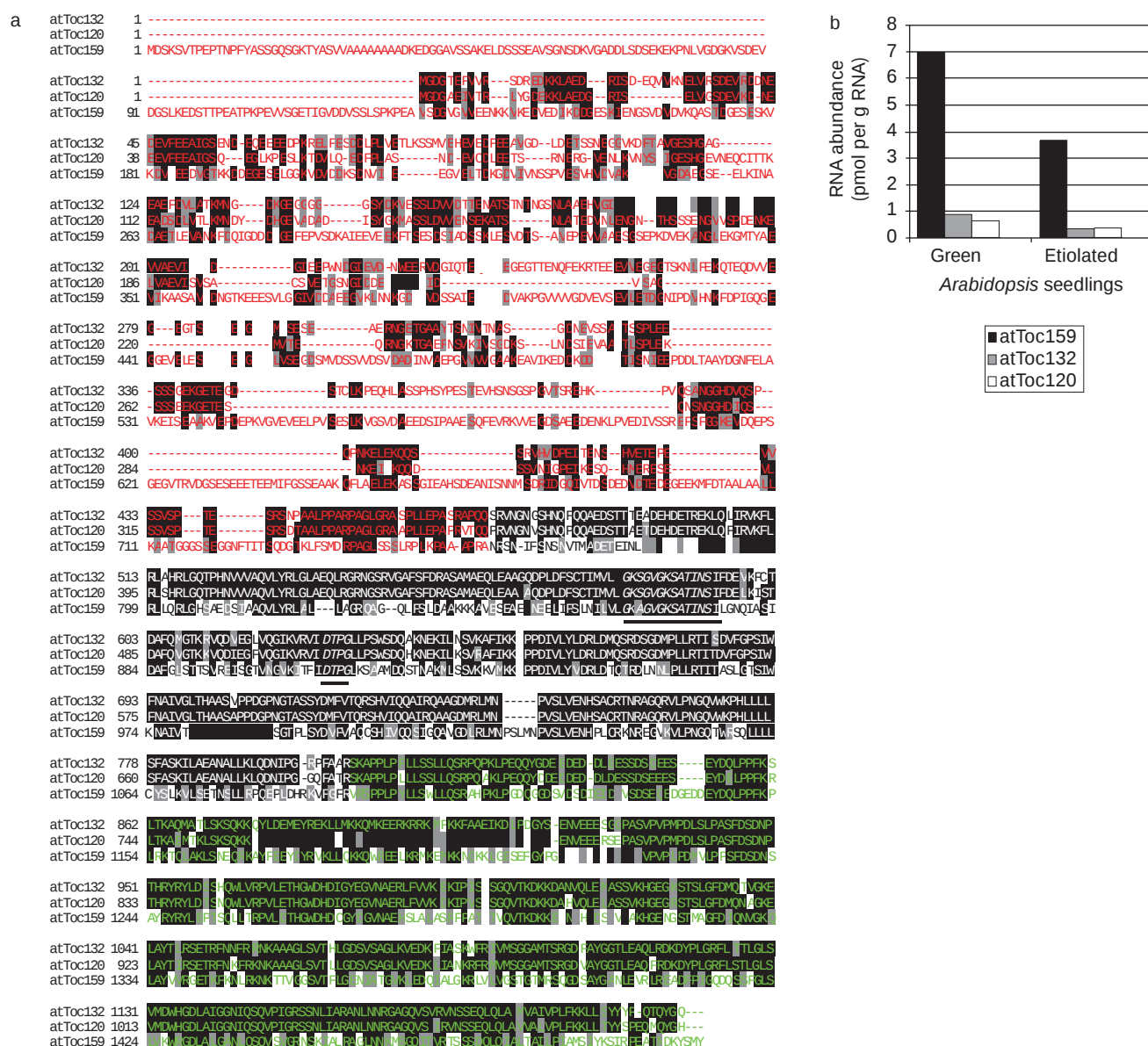


Figure 1 A family of three Toc159-related genes in *Arabidopsis*. **a**, Deduced sequences of atToc159, atToc132 and atToc120 were aligned using the ClustalW 1.7 software. Gaps are introduced to maximize identical sequences. Amino acids identical in at least two of the sequences are shaded in black; conserved substitutions are shaded in grey. Variable

acidic domains are in red, the GTPase region is in black and white, and the membrane anchor domain is in green. GTP-binding motives are underlined. **b**, Abundance of atToc159, -132 and -120 RNA in *Arabidopsis* seedlings grown in continuous light (Green) or in the dark (Etiolated) for 6 days.

all three proteins are increased about twofold in light-grown green as opposed to dark-grown etiolated seedlings (Fig. 1b). This finding presumably reflects increased import activity in developing chloroplasts when expression levels of nuclear-encoded photosynthetic proteins are highest^{17,18}.

Synthetic [³⁵S]atToc159, -132 and -120 (Fig. 2a, lane 1) were inserted into the outer membrane of isolated pea chloroplasts in an *in vitro* import reaction (Fig. 2a, lane 2) to verify that the putative receptor proteins are chloroplast proteins. All three proteins were resistant to alkaline extraction (Fig. 2a, lane 4). Thermolysin treatment of chloroplasts after the import reactions resulted in the quantitative degradation of the imported proteins to membrane anchor fragments of relative molecular mass 52,000 (*M_r* 52K; Fig. 2a, lane 3), as previously shown for psToc159 (ref. 5). These results confirm that the proteins are found in the outer envelope membrane, and indicate that all three integrate into the membrane with their N-terminal acidic and GTP-binding domains oriented to the cytoplasm. The location and topology of endogenous atToc159 was confirmed by thermolysin treatment of isolated *Arabidopsis* chloroplasts followed by immunoblotting with anti-atToc159₁₋₇₄₀ antibodies (data not shown). Portions of newly imported [³⁵S]atToc159, -132 or -120 associated with Toc complexes that were immunoaffinity-purified using anti-psToc34 IgG-Sepharose (Fig. 2b, lane 2). No proteins were detected in control experiments using preimmune IgG-Sepharose (Fig. 2b, lane 3). These data lead us to conclude that atToc159, -132 and -120 function as components of the Toc complex.

The divergence in the cytoplasmic acidic domains among atToc159, -132 and -120 indicates that they have distinct roles in protein import. To determine the role of atToc159, the most abundant member of the group, we screened T-DNA mutant collections¹⁹ for insertions in the *TOC159* gene. Two independent mutant lines, CS11072 and CS19917, were obtained (Fig. 3a) that contain insertions in *TOC159* at 770 and ~1,600 base pairs (bp),

respectively. The mutant lines have identical albino phenotypes (Fig. 3b). CS11072, termed *ppi2*, contains a single T-DNA insertion (data not shown) and so was selected for further characterization. Polymerase chain reaction with reverse transcription (RT-PCR) and immunoblotting analyses confirmed the absence of Toc159 mRNA and protein in extracts of CS11072, indicating that *ppi2* is a null mutant (data not shown). Mutant plants are not viable on soil beyond the cotyledon stage (Fig. 3b), so *ppi2* is seedling lethal. When viewed by transmission electron microscopy, *ppi2* plastids lack the thylakoid membranes and starch granules of mature chloroplasts (Fig. 3d), and therefore resembled undifferentiated proplastids (Fig. 3c). Furthermore, *ppi2* plastids developed only a rudimentary paracrystalline internal membrane structure (prolamellar body) that is characteristic of etioplasts when grown in the dark on sucrose (data not shown). These results indicate that the ability of *ppi2* proplastids to differentiate into chloroplasts is blocked, but other plastid functions, such as etioplast formation, might also be affected.

Coomassie blue (Fig. 4a) and immunoblot (Fig. 4b) analyses revealed that three major photosynthetic genes, the large (RbcL) and small (RbcS) subunits of Rubisco (a known import substrate of Toc159 in pea⁵⁻⁷) and the chlorophyll *a/b* binding protein (Cab) are present in *ppi2* at much lower levels than in wild-type plants. Furthermore, the expression of all three proteins was strongly reduced in *ppi2* plants (Fig. 5a). These data show that *ppi2* results in transcriptional repression of highly expressed photosynthetic genes. The simultaneous repression of RbcS and RbcL, a protein encoded by the chloroplast genome, is expected as their regulation is coupled²⁰. In contrast, the expression of the light-induced cytoplasmic enzyme chalcone synthase²¹ was not reduced in *ppi2* (Fig. 5a), indicating that the repression of transcription is specific for the genes encoding chloroplast proteins. The selective effect of *ppi2* on the transcription of photosynthetic genes is typical of *Arabidopsis* mutants that are defective in chloroplast biogenesis²². Although

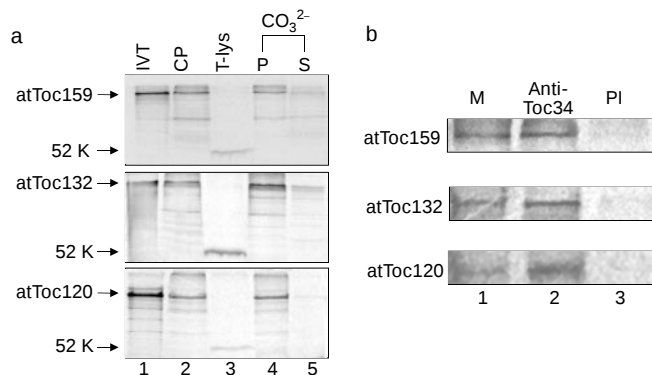


Figure 2 atToc159, atToc132 and atToc120 are components of the Toc complex of the plastid protein import machinery. **a**, Synthetic [³⁵S]atToc159, -132 or -120 (IVT) associate with isolated pea chloroplasts in a standard import reaction (CP). Thermolysin treatment (T-lys) degrades all three imported proteins to membrane-protected fragments of relative molecular mass 52,000 (52K). Imported [³⁵S]atToc159, -132 and -120 remain with the membrane fraction (P) after extraction with alkaline carbonate buffer (CO₃²⁻). **b**, Imported [³⁵S]atToc159, -132 or -120 associated with the Toc complex. They indirectly bind to anti-Toc34 IgG-Sepharose but not preimmune IgG-Sepharose (PI), indicating that they associate with Toc complexes. Lane 1 contains 10% of the chloroplast membrane fraction (M) used for immunoprecipitations.

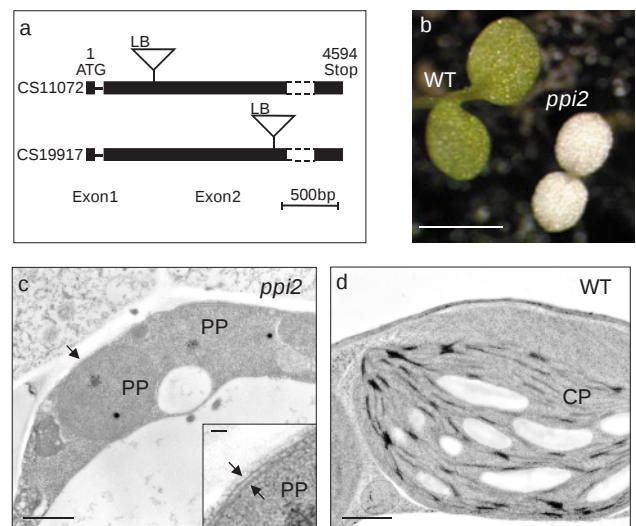


Figure 3 Characteristics of the *ppi2* mutant. **a**, Schematic representation of *TOC159* gene disruptions in the CS11072 and CS19917 *Arabidopsis* lines. Translation initiation (ATG) and termination (Stop) codons are indicated. The T-DNA inserts are represented with the 5' border sequences of the insert (LB) labelled to indicate the orientation (not to scale). **b**, Visible phenotype of *ppi2*. CS11072 *ppi2* line and wild-type (WT) seedlings were grown in soil for 6 days in long-day conditions (16 hours light: 8 hours dark). Scale bar, 2 mm. **c**, **d**, Ultrastructure of *ppi2* plastids. Transmission electron microscopy of plants grown on soil for 6 days in long-day conditions indicates that *ppi2* plastids (**c**) remain as undifferentiated proplastids (PP) compared to the chloroplasts (CP) present in wild-type (WT) cells (**d**). Scale bar, 0.5 μm. Inset, the double membrane envelope of the *ppi2* proplastids; scale bar, 0.05 μm.

RbcS and Cab expression was reduced, the proteins appeared to be normally processed in *ppi2* because their electrophoretic mobilities were identical to those in wild-type plants (Fig. 4b). Indeed, RbcS (Fig. 4c) and Cab (data not shown), as in wild-type plants (Fig. 4d), are located exclusively within plastids, rather than accumulated in the cytosol in *ppi2*, as judged by immunoelectron microscopy. Thus, although atToc159 is essential for chloroplast biogenesis, plastids lacking this Toc component are still capable of protein import.

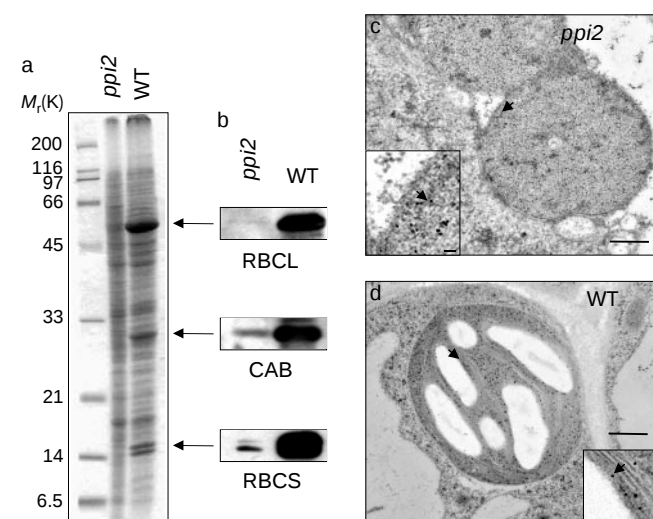


Figure 4 *ppi2* plants contain reduced amounts of photosynthetic proteins. **a**, The reduction in RbcS, RbCL and Cab expression is apparent by SDS-PAGE analysis of wild-type (WT) and *ppi2* plants. **b**, Immunoblotting with sera against the large (RBCL) and small (RBCS) subunits of Rubisco and the chlorophyll *a/b* binding protein (CAB) reveals a dramatic reduction of the proteins in *ppi2* plants. RbcS and Cab are processed to their mature forms in the *ppi2* plants. **c, d**, Immunogold labelling of RbcS in plastids of *ppi2* (**c**) and wild-type (**d**) cotyledons, indicating that RbcS is imported in the absence of atToc159. Scale bar, 0.5 μ m. Arrows indicate the position of the insets. The arrow in the inset points to a gold particle; scale bar, 0.05 μ m.

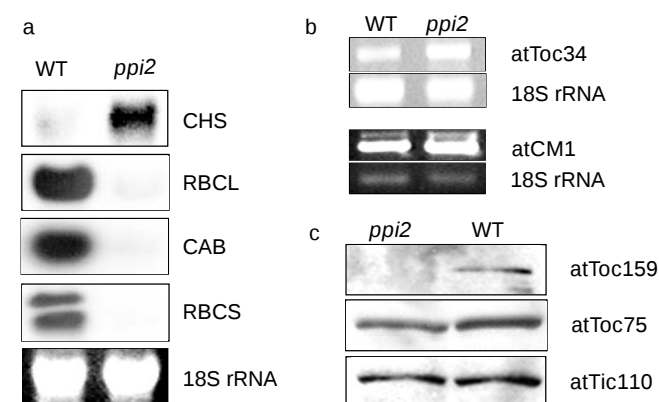


Figure 5 Characterization of the chloroplast biogenetic defect in *ppi2* plants. **a**, Comparative northern-blot analysis of light-induced mRNAs from wild-type (WT) and *ppi2* 6-day-old seedlings. RbCL, RbcS and Cab transcription is repressed in the *ppi2* mutant, whereas the light-regulated gene of cytoplasmic chalcone synthase (CHS) is not repressed. 18S ribosomal RNA was used to normalize loading. **b**, Comparative RT-PCR analysis of non-photosynthetic genes from wild-type and *ppi2* plants. The amount of mRNA encoding chorismate mutase 1 (atCM1) is unchanged in *ppi2* plants. atToc34 mRNA levels are upregulated in the *ppi2* mutant. **c**, Immunoblot analysis of atTic110 and atToc75 in protein extracts from wild-type and *ppi2* plants. The expression and processing of atTic110 and atToc75 are normal in mutant plants, indicating that these proteins are imported in the absence of atToc159.

In contrast to RbcS and Cab, the expression of chorismate mutase (atCM1), a plastid protein that is not specific to chloroplasts²³, was unaffected by *ppi2* (Fig. 5b). The expression of the gene encoding atToc34 (ref. 14) was increased in *ppi2*, possibly to compensate for the loss of atToc159. Furthermore, immunoblots of *ppi2* and wild-type extracts show that expression and processing of atToc75 and atTic110 (refs 8, 9), a component of the Tic-complex, were comparable (Fig. 5c). Toc75 and Tic110 are targeted to chloroplasts by N-terminal transit sequences through the Toc complex^{18,24}. The expression and import of the non-photosynthetic proteins tested is therefore not affected in *ppi2*. The presence of atToc34, atToc75 and atTic110 in *ppi2* plants provides additional evidence that the import apparatus is expressed and functional in the absence of atToc159. A likely explanation for this observation is that atToc132 and/or atToc120, which are expressed in *ppi2* (data not shown), partly compensate for the absence of atToc159.

The characteristics of the *ppi2* mutant demonstrate that protein import is a limiting process in chloroplast biogenesis and reveal alternative pathways for targeting proteins to plastids. The atToc159 defect limits the capacity of plastids to import a set of highly expressed photosynthetic proteins that are essential for chloroplast biogenesis. atToc132 and atToc120 might represent additional import complexes with distinct but partly overlapping substrate specificities, which might account for the ability of *ppi2* plastids to import other proteins. The functional differences among the import receptors are likely to involve the acidic cytoplasmic domains where atToc159, -132 and -120 are most divergent. □

Methods

Isolation of complementary DNAs and production of antisera

A cDNA encompassing the coding region of the atToc159 gene was amplified directly from *Arabidopsis* genomic DNA by PCR. The 5' PCR primer (CAT GCC ATG GAC TCA AAG TCG GTT ACT CCA GAA CCA ACC AAC CCC TTC TAC GCT TCT TCG GGG CAA TCA GGA AAA ACC TAT GCT TCT GTT GTC) incorporated a 5' *Nco*I site and encoded the entire first exon and 26 bases of the 5' end of the second exon of the atToc159 gene (Genbank accession no. AC002330). The 3' primer corresponded to the 3' end of the gene and incorporated a 3' *Sac*I site. The resulting atToc159 cDNA was ligated into pET21d (Novagen). The atToc132 (Genbank accession no. AC005825) and atToc120 (Genbank accession no. AB02217) cDNAs were amplified from total RNA of 10-day-old *Arabidopsis* seedlings using primers corresponding exactly to the 5' and 3' ends of the coding regions. atToc132 and atToc120 cDNAs were ligated into pCR-XL-TOPO (Invitrogen). Anti-atToc159 and anti-atTic110 antibodies were raised against *Escherichia coli*-expressed N-terminal portions of atToc159 (amino acids 1 to 740) and atTic110 (amino acids 1 to 498), respectively. The antibodies were affinity purified against the respective antigen before use.

Isolation of T-DNA insertion mutants

PCR-based identification of T-DNA insertions in *TOC159* was performed according to published protocols¹⁵. DNA pools from the *Arabidopsis* libraries of T-DNA insertion mutants were obtained from the *Arabidopsis* Biological Resource Center, Ohio State University, Columbus, Ohio. PCR reactions required 29-mer primers corresponding to the 5' end (forward primer, ATG GAC TCA AAG TCG GTT ACT CCA GAA CC) and 3' end (reverse primer, TTA GTA CAT GCT GTA CTT GTC GTT CGT CG) of atToc159 and the left and right border of the T-DNA¹⁵, respectively. Line CS11072 carried a single insertion consisting of two concatamers right-border to right-border T-DNA molecules, as indicated by Southern blot and PCR analysis and the 3:1 segregation of the T-DNA kanamycin-resistance marker.

Chloroplast import experiments

[³⁵S]atToc159, -132 and -120 were synthesized in a coupled *in vitro* transcription and translation system in the presence of [³⁵S]methionine (Promega). Protein import into isolated pea chloroplasts, thermolysin treatment, chloroplast-membrane isolation and carbonate extraction were performed as described²⁵. Immunoaffinity chromatography using anti-Toc34 IgG-Sepharose was performed as described²⁶.

Characterization of the *ppi2* mutation

Wild-type and mutant plants were grown for 6 days either on soil under long-day conditions (16 hours light: 8 hours dark) or on agarose plates containing 0.5× Murashige-Skoog medium and 1% sucrose (MS-medium) in continuous light. Protein extracts for SDS-PAGE and immunoblot analyses were prepared by direct extraction of seedling tissue in boiling SDS-PAGE sample buffer. Samples corresponding to equivalent amounts of fresh mass were loaded onto SDS-PAGE gels. Antisera were raised against *Arabidopsis* chlorophyll *a/b* binding protein, pea large subunit of Rubisco, *Chlamydomonas reinhardtii* small subunit of Rubisco, and pea Toc75. Northern blots containing 5 μ g of total RNA from 6-day-old *Arabidopsis* plants grown on MS medium were hybridized to random-primed

[³²P]cDNA or genomic DNA probes. The probes were generated by PCR amplification of chalcone synthase (Swiss-Prot accession no. P13114, base pairs 346–1082), large and small subunits of Rubisco (Genbank accession no. U91966 (base pairs 1–1481) and g16195 (base pairs 1–552)), and the chlorophyll *a/b* binding protein (Genbank accession no. X03908 (base pairs 507–999)) of *Arabidopsis*. Relative competitive RT-PCR analysis was performed according to the manufacturer's protocols (Ambion). Specific primer pairs to the following genes were used: atToc34 (expressed sequence tag clone 190117T7: forward, ATG GCC ATG GGG TCT CTC GTG CGT GAA TGG; reverse, TGC GGA TCC TTA AAG TGG CTT TCC ACT TGT); atCM1 (Genbank accession no. Z26519: forward, ATG AGA TCG TCT TGT TGC TCC; reverse, TCA GTC CAG TCT TCT GAG CAA G); atToc159 (forward, CAC AGT CTT GCT CTA GCT AGC CGG TTC; reverse, GCT GTA CTT GTC GTT CGT CGC TTC); atToc132 (forward, GAT TCG GTT TCT GCG GGG TTG; reverse, TCA TTG TCC ATATTG CGT TTG CGG); atToc120 (forward, AAT GCT GGG AAG GAA TTA GCG TAC ACT A; reverse, TCA GTG TCC ATA TTG CAT TTG CTC AGG). Electron microscopy and immunoelectron microscopy were performed according to previously published protocols²⁷.

Received 17 August; accepted 21 October 1999.

- Kendrick, R. E. & Kronenberg, G. H. M. *Photomorphogenesis in Plants* (Junk, Dordrecht, 1994).
- Chen, X. & Schnell, D. J. Protein import into chloroplasts. *Trends Cell Biol.* **9**, 222–227 (1999).
- Keegstra, K. & Cline, K. Protein import and routing systems of chloroplasts. *Plant Cell* **11**, 557–570 (1999).
- Soll, J. & Tien, R. Protein translocation into and across the chloroplastic envelope membranes. *Plant Mol. Biol.* **38**, 191–207 (1998).
- Hirsch, S., Muckel, E., Heemeyer, F., von Heijne, G. & Soll, J. A receptor component of the chloroplast protein translocation machinery. *Science* **266**, 1989–1992 (1994).
- Schnell, D. J., Kessler, F. & Blobel, G. Isolation of components of the chloroplast protein import machinery. *Science* **266**, 1007–1012 (1994).
- Perry, S. E. & Keegstra, K. Envelope membrane proteins that interact with chloroplastic precursor proteins. *Plant Cell* **6**, 93–105 (1994).
- Kessler, F. & Blobel, G. Interaction of the protein import and folding machineries in the chloroplast. *Proc. Natl Acad. Sci. USA* **93**, 7684–7689 (1996).
- Lübeck, J., Soll, J., Akita, M., Nielsen, E. & Keegstra, K. Topology of IEP110, a component of the chloroplastic protein import machinery present in the inner envelope membrane. *EMBO J.* **15**, 4230–4238 (1996).
- Chen, K., Chen, X. & Schnell, D. J. Initial binding of preproteins involving the Toc159 receptor can be bypassed during protein import into chloroplasts. *Plant Phys.* (in the press).
- Seedorf, M., Waagemann, K. & Soll, J. A constituent of the chloroplast import complex represents a new type of GTP-binding protein. *Plant J.* **7**, 401–411 (1995).
- Kessler, F., Blobel, G., Patel, H. A. & Schnell, D. J. Identification of two GTP-binding proteins in the chloroplast protein import machinery. *Science* **266**, 1035–1039 (1994).
- Ma, Y., Kouranov, A., LaSala, S. E. & Schnell, D. J. Two components of the chloroplast protein import apparatus, IAP86 and IAP75, interact with the transit sequence during the recognition and translocation of precursor proteins at the outer envelope. *J. Cell Biol.* **134**, 1–13 (1996).
- Jarvis, P. et al. An *Arabidopsis* mutant defective in the plastid general protein import apparatus. *Science* **282**, 100–103 (1998).
- McKinney, E. C. et al. Sequence-based identification of T-DNA insertion mutations in *Arabidopsis*: actin mutants act2-1 and act4-1. *Plant J.* **8**, 613–622 (1995).
- Bölter, B., May, T. & Soll, J. A protein import receptor in pea chloroplasts, Toc86, is only a proteolytic fragment of a larger polypeptide. *FEBS Lett.* **441**, 59–62 (1998).
- Dahlin, C. & Cline, K. Developmental regulation of the plastid protein import apparatus. *Plant Cell* **3**, 1131–1140 (1991).
- Tranel, P. J., Froehlich, J., Goyal, A. & Keegstra, K. A component of the chloroplastic protein import apparatus is targeted to the outer envelope membrane via a novel pathway. *EMBO J.* **14**, 2436–2446 (1995).
- Feldmann, K. T-DNA insertion mutagenesis in *Arabidopsis*—mutational spectrum. *Plant J.* **1**, 71–82 (1991).
- Surpin, M. & Chory, J. The co-ordination of nuclear and organellar genome expression in eukaryotic cells. *Essays Biochem.* **32**, 113–125 (1997).
- Ang, L.-H. et al. Molecular interaction between COP1 and HY5 defines a regulatory switch for light control of *Arabidopsis* development. *Mol. Cell* **1**, 213–222 (1998).
- Li, H.-M., Culligan, K., Dixon, R. A. & Chory, J. CUE1: A mesophyll cell-specific positive regulator of light-controlled gene expression in *Arabidopsis*. *Plant Cell* **7**, 1599–1610 (1995).
- Eberhard, J. et al. Cytosolic and plastidic chorismate mutase isozymes from *Arabidopsis thaliana*: molecular characterization and enzymatic properties. *Plant J.* **10**, 815–821 (1996).
- Lübeck, J., Heins, L. & Soll, J. A nuclear encoded chloroplastic inner envelope membrane protein uses a soluble sorting intermediate upon import into the organelle. *J. Cell Biol.* **137**, 1279–1286 (1997).
- Chen, D. & Schnell, D. J. Insertion of the 34 kDa chloroplast protein import component, IAP34, into the chloroplast outer membrane is dependent on its intrinsic GTP-binding capacity. *J. Biol. Chem.* **272**, 6614–6620 (1997).
- Kouranov, A., Chen, X., Fuks, B. & Schnell, D. J. Tic20 and Tic22 are new components of the protein import apparatus at the chloroplast inner envelope membrane. *J. Cell Biol.* **143**, 991–1002 (1998).
- Michel, M., Hillmann, T. & Müller, M. Cryosectioning of plant material frozen at high pressure. *J. Microsc.* **163**, 3–18 (1991).

Acknowledgements

We thank G. Armstrong, Q. Su and K. Apel for antibodies, cDNA probes and discussion; N. Amrhein, P. Macheroux and A. Schaller for support and discussion; and G. Schatz for encouragement. D.S. was supported by grants from the National Science Foundation and a Charles and Johanna Busch Memorial Fund award. F.K. was supported by grants from the Swiss Federal Institute of Technology and the Swiss National Science Foundation.

Correspondence and requests for materials should be addressed to F.K. (e-mail: felix.kessler@ipw.biol.ethz.ch) or D.J.S. (e-mail: schnell@andromeda.rutgers.edu).

Interferon- γ elicits arteriosclerosis in the absence of leukocytes

George Tellides^{*†}, Denis A. Tereb^{*†}, Nancy C. Kirkiles-Smith^{*†}, Richard W. Kim^{*†}, Jean H. Wilson[†], Jeffrey S. Schechner^{*‡}, Marc I. Lorber[†] & Jordan S. Pober^{*§||}

^{*} Interdepartmental Program in Vascular Biology and Transplantation, Boyer Center for Molecular Medicine, and the Departments of [†] Surgery, [‡] Dermatology, [§] Pathology and ^{||} Immunobiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

Atherosclerosis and post-transplant graft arteriosclerosis are both characterized by expansion of the arterial intima as a result of the infiltration of mononuclear leukocytes, the proliferation of vascular smooth muscle cells (VSMCs) and the accumulation of extracellular matrix^{1–3}. They are also associated with the presence of the immunomodulatory cytokine interferon- γ (IFN- γ)^{2,3}. Moreover, in mouse models of atheroma formation or allogeneic transplantation, the serological neutralization⁴ or genetic absence^{5–8} of IFN- γ markedly reduces the extent of intimal expansion. However, other studies have found that exogenous IFN- γ inhibits cultured VSMC proliferation^{9–14} and matrix synthesis¹⁵, and reduces intimal expansion in response to mechanical injury^{16–18}. This discrepancy is generally explained by the idea that IFN- γ either directly activates macrophages, or, by increasing antigen presentation, indirectly activates T cells within the lesions of atherosclerosis and graft arteriosclerosis. These activated leukocytes are thought to express the VSMC-activating cytokines^{1–3}

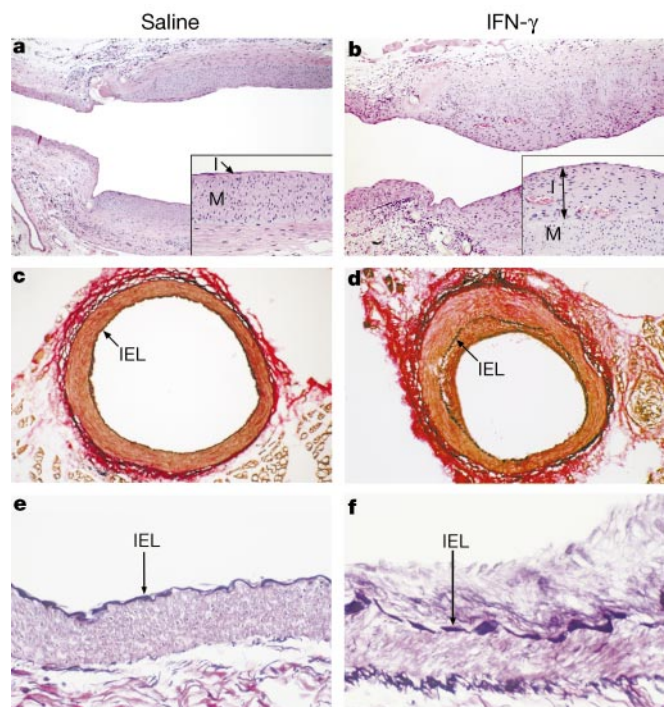


Figure 1 IFN- γ elicited arteriosclerosis *in vivo*. Histological appearance of pig coronary (a–d) and human internal mammary (e, f) artery grafts 35 days after transplantation to SCID/beige mice treated with saline (a, c, e), pIFN- γ (b, d) or hIFN- γ (f). Representative photomicrographs show H&E-stained longitudinal sections at low magnification with high-magnification insets (a, b), EVG-stained cross-sections at low magnification (c, d), and EVG-stained longitudinal sections at high magnification (e, f). The intima (I), media (M) and internal elastic lamina (IEL) are labelled. Sutures between mouse aorta and pig coronary artery are visible as clear material.

and cell-surface molecules¹⁹ that cause the observed arteriosclerotic responses. Here we have inserted pig and human arteries into the aorta of immunodeficient mice, and we show that IFN- γ can induce arteriosclerotic changes in the absence of detectable immunocytes by acting on VSMCs to potentiate growth-factor-induced mitogenesis.

To test the hypothesis that IFN- γ directly elicits arteriosclerosis, we developed a xenotransplantation model to exploit the species specificity of this cytokine. Pig coronary arteries were interposed into the aorta of severe combined immunodeficient (SCID)/beige mice. Pigs were selected as donors because of their availability and the close resemblance of swine and human coronary atherosclerosis. We repeated some experiments using human arteries to exclude the possibility of a species-restricted effect. SCID/beige mice were chosen as hosts because deficiencies of their T and B lymphocytes and natural killer cell function mean they cannot reject xenografts or develop antibodies to xenogeneic cytokines²⁰. In these chimaeric subjects, exogenous porcine and human IFN- γ (pIFN- γ and hIFN- γ , respectively) can act only on cells of donor origin and not on mouse cells²⁰.

Initial experiments demonstrated that pig coronary arteries from saline-treated SCID/beige mice have a normal histological appear-

ance at 1, 2, 3, 7, 14, 21, 28, 35 and 70 days post-operatively ($n = 3-6$ at each time point). The endothelium was preserved without adherent thrombi, the intima was not thickened, the media was viable without atrophy, and the adventitia was not fibrosed, although mouse granulation tissue formed around the grafts (Fig. 1a, c). Similarly, human arteries did not acquire pathological changes within 35 days of transplantation ($n = 6$; Fig. 1e).

We then investigated the effects of IFN- γ on arterial morphology. In contrast to control grafts, pig coronary arteries from pIFN- γ -treated animals had a mild thickening of the intima after 28 days ($n = 3$; data not shown) and a moderate to marked expansion of the intima by day 35 ($n = 6$; Fig. 1b, d). The neointima was asymmetric with a matrix-rich appearance and ample neovascularization, and the internal elastic lamina remained intact. The adjacent mouse aorta appeared normal (Fig. 1a, b) and there was no recipient coronary arteriosclerosis (data not shown), consistent with the species-specific effects of IFN- γ . These results also excluded a systemic response of the host responsible for the arteriosclerotic process, such as the altered hepatic lipoprotein metabolism seen in mice deficient in the IFN- γ receptor⁵. Human coronary, inferior epigastric and internal mammary arteries that were transplanted to hIFN- γ -treated SCID/beige mice also developed intimal expansion by day 35 ($n = 9$; Fig. 1f), so the vascular response to IFN- γ was not restricted to a particular species or organ.

The graft phenotype was determined by immunohistochemistry. Pig major histocompatibility complex (MHC) class I and II molecules were strongly expressed by endothelial cells before and one day after transplantation, were focally and weakly present after 7 days, and were absent 14 and 35 days post-operatively in saline-treated animals ($n = 3-6$ at each time point; Fig. 2a, b, d, e). In contrast, pIFN- γ administration sustained porcine endothelial MHC class I and II antigen expression at pretransplant levels up to day 35 (Fig. 2c, f). Similarly, pIFN- γ restored the weak basal expression of pig MHC class I antigen by graft VSMCs (Fig. 2c). Porcine platelet-endothelial cell adhesion molecule-1 (PECAM-1) was constitutively expressed by pig endothelium in both control and cytokine-treated mice (Fig. 2g-i). Mouse MHC class I antigen was weakly expressed and mouse PECAM-1 was strongly expressed by mouse aortic endothelial cells (data not shown). Murine antigens were present in the recipient granulation tissue, but not in the pig coronary arteries (Fig. 2a-c, g-i, insets). The immunostaining results established that donor endothelial cells and VSMCs are retained by the graft and are not replaced by cells of host origin. Lack of mouse PECAM-1 staining on the pig endothelium also excluded the adherence of mouse platelets, which can contribute to arteriosclerosis¹.

The cells in the graft neointima after pIFN- γ treatment stained positive for smooth muscle α -actin, as did the cells of the media, identifying them as VSMCs (Fig. 2j-l). Pig passenger leukocytes,

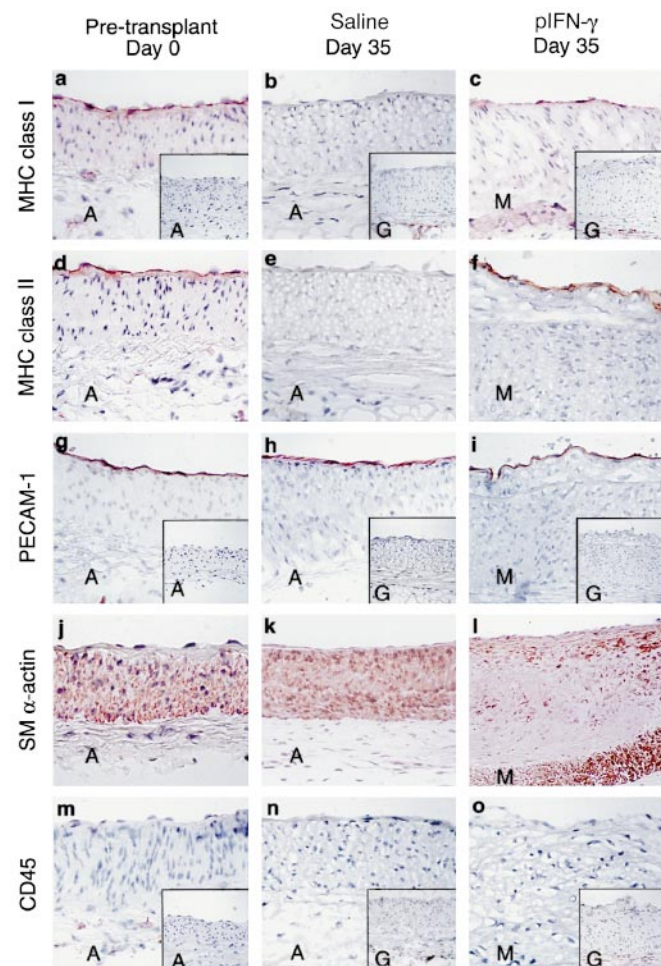


Figure 2 IFN- γ sustained endothelial MHC antigen expression and resulted in an accumulation of intimal VSMCs in the absence of leukocytes. Immunohistochemical analysis determined the cellular phenotype of pig coronary artery grafts before and 35 days after transplantation to saline- or pIFN- γ -treated SCID/beige mice using monoclonal antibodies to pig MHC class I antigen (a-c), MHC class II antigen (d-f), PECAM-1 (g-i), smooth muscle (SM) α -actin (j-l), CD45 (m-o) and corresponding mouse antigens (insets). Representative photomicrographs (high magnification, except for l) of longitudinal graft sections are orientated with the lumen above and the media (M), adventitia (A) or mouse granulation tissue (G) below.

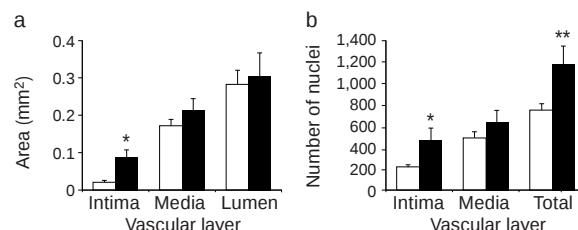


Figure 3 IFN- γ increased the intima area and number of intimal nuclei. SCID/beige mouse recipients of pig coronary artery grafts were treated with heat-inactivated cytokine (open bars, $n = 6$) or pIFN- γ (filled bars, $n = 7$) from post-operative day 7 to day 35. **a**, Cross-sectional areas of intima, media and lumen were quantified by computer-assisted image analysis. **b**, The nuclei in cross-sections of the vessel wall were counted using microscopy. Data are mean $\pm$ s.e.m. Statistical significance was determined by the Mann-Whitney U test. * $P < 0.01$, ** $P < 0.05$.

which were present in the adventitia before transplantation, had greatly diminished in number after the first day, and were undetectable after 7, 14 and 35 days in grafts from both saline- and cytokine-treated animals (Fig. 2m–o). Mouse leukocytes were found in the host granulation tissue around the pig arteries (Fig. 2n–o, insets), were rarely adherent to the graft endothelium, but did not infiltrate the vascular wall under any treatment. Thus, within the detection limits of immunohistochemistry, IFN γ -mediated arteriosclerosis seems to be independent of donor and host leukocytes.

We used computer-assisted image analysis to quantify the degree of arteriosclerosis. pIFN γ treatment resulted in a marked enlargement of the intima cross-sectional area, but not of the media or lumen areas, compared with heat-inactivated cytokine treatment (Fig. 3a). The abluminal expansion of the intima did not compromise the lumen as there was a compensatory distention of the artery, a phenomenon of early atherosclerosis known as remodelling¹. A similar exuberant neointima developed in pig coronary arteries from a separate group of animals, whose vascular grafts were processed as longitudinal sections for histology, where the maximal intima thickness in pIFN γ -treated recipients ($n = 4$) was $91 \pm 7 \mu\text{m}$, compared with $7 \pm 4 \mu\text{m}$ ($P < 0.05$) in saline-treated controls ($n = 4$). Human arteries from hIFN γ -treated SCID/beige mice ($n = 9$) also developed an expanded intima of $126 \pm 41 \mu\text{m}$ compared with $24 \pm 12 \mu\text{m}$ ($P < 0.05$) in untreated recipients ($n = 6$).

The number of graft nuclei was quantified by microscopy. Administration of pIFN γ resulted in a substantial increase in the number of intimal and total vascular nuclei, compared with heat-inactivated cytokine treatment (Fig. 3b). Changes in the media were not significant. The accumulation of VSMCs within the arterial intima, without loss of medial VSMCs, cannot be explained by cell migration across the internal elastic lamina alone and must be due to increased proliferation and/or survival of these cells. Cell division within the transplanted arteries was supported by the expression of proliferating cell nuclear antigen (PCNA) and the uptake of bromodeoxyuridine (BrdU) by intimal and, to a lesser degree, by medial VSMCs from pIFN γ -treated, but not saline-treated, animals (Fig. 4a–d).

To analyse further the basis of VSMC proliferation, we examined the grafts for growth factors and their receptors. Immunohistochemical analysis revealed increased expression of platelet-derived growth factor (PDGF) and PDGF β -receptors in medial and, especially, intimal VSMCs of pig coronary arteries from animals treated with pIFN γ , compared with saline (Fig. 4e–h). In contrast, expression of transforming growth factor- β (TGF- β) and TGF- β type-II receptors was not altered by pIFN γ treatment (Fig. 4i–l). Similar immunohistochemical findings were found in human coronary arteries in response to hIFN γ (data not shown). Induction of PDGF and PDGF β -receptors has been previously described in human atherosclerotic plaques and is thought to play a role in the formation of the lesions^{2,3}.

We then investigated whether IFN γ had a mitogenic effect on cultured cells. Pig coronary artery VSMC proliferation was unaffected by pIFN γ under either low or optimal serum conditions for up to 7 days (Fig. 5a). Flow-cytometric DNA-content analysis verified that pIFN γ did not affect the rate of VSMC replication (the fraction of cells with supradiploid DNA content) or apoptosis (the fraction of cells with subdiploid DNA content) after 1 (Fig. 5b), 3, 5 and 7 days (data not shown). Similarly, tenfold higher and lower concentrations of pIFN γ had no significant effect on VSMC division, even though MHC antigens were induced in a dose- and time-dependent fashion (data not shown).

Although IFN γ was not found to be directly mitogenic, further experiments showed that pIFN γ potentiated the proliferative effect of porcine PDGF-BB, but not that of TGF- β 1 (data not shown) or interleukin-1 β (IL-1 β), and the growth augmentation of quiescent VSMCs was significantly greater with 10% than 0.5% serum (Fig. 5c). Preincubation of VSMCs with pIFN γ under low-serum conditions markedly potentiated subsequent PDGF-induced cell division (Fig. 5d). In contrast, pIFN γ failed to stimulate and sometimes suppressed the multiplication of exponentially dividing VSMCs or that of cells undergoing apoptosis in the absence of serum (our unpublished observations). These condition-dependent responses may have confounded some previous reports on the effect of IFN γ on VSMC growth^{9–14}.

Finally, because cytokine pretreatment enhance later growth-factor effects, we examined whether IFN γ affected the expression

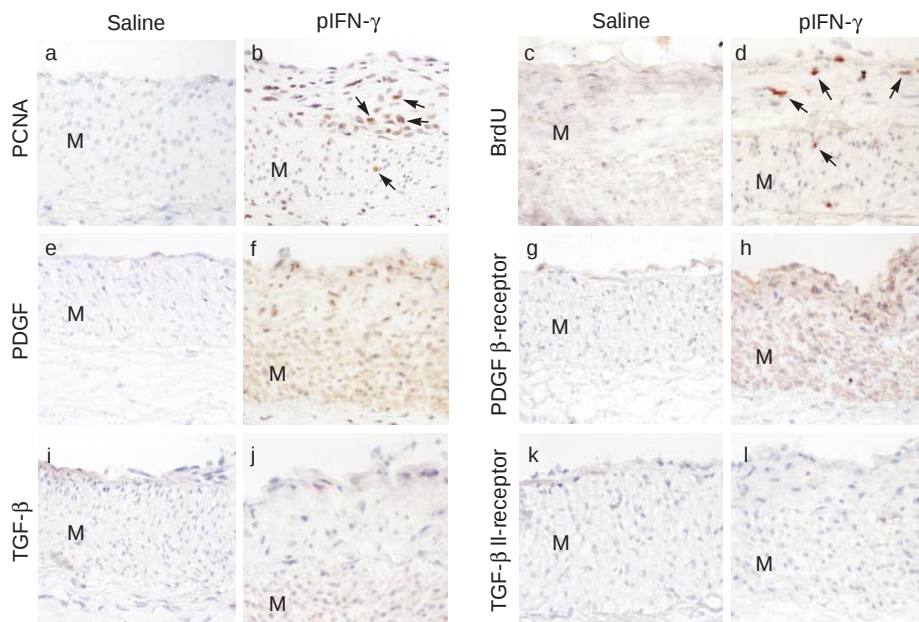


Figure 4 IFN γ increased VSMC proliferation and expression of PDGF and its receptor. Immunohistochemical analysis of pig coronary artery grafts from saline- or pIFN γ -treated SCID/beige mice at post-operative day 35 using antibodies to PCNA (**a, b**), BrdU (**c, d**), PDGF-AB (**e, f**), PDGF β -receptor (**g, h**), TGF- β (**i, j**) and TGF- β type-II receptor

(**k, l**). Representative photomicrographs of longitudinal graft sections at high magnification are orientated with the lumen above. Proliferating cells (arrows) and the media (M) are labelled.

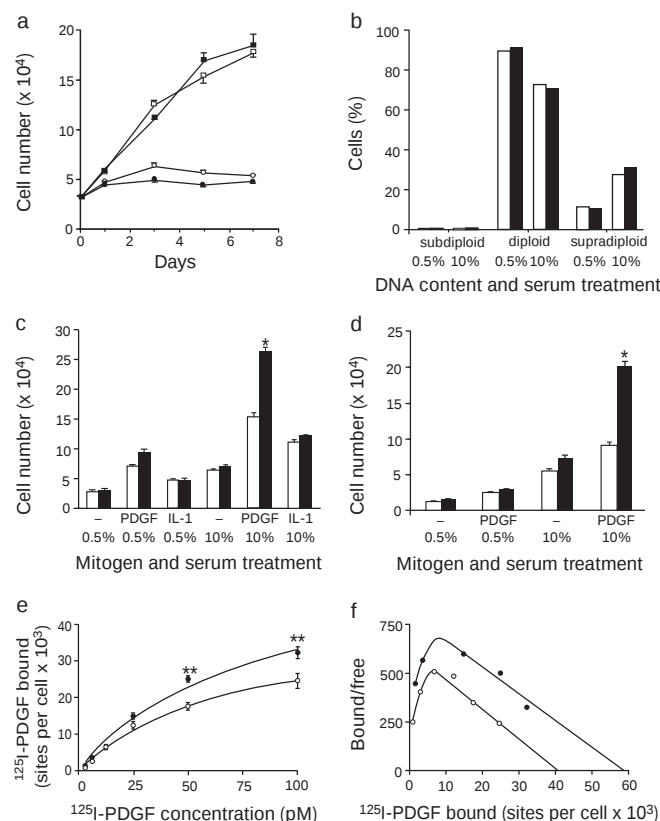


Figure 5 IFN- γ was not a direct mitogen, but potentiated PDGF-induced proliferation and upregulated PDGF receptor expression *in vitro*. Pig coronary artery VSMCs were cultured in the presence (filled bars or symbols) or absence (open bars or symbols) of pIFN- γ (10 ng ml⁻¹). **a**, VSMCs were grown in 0.5% (circles) or 10% (squares) serum with or without pIFN- γ and counted at 1–7 days. **b**, Their DNA content was quantified after 1 day. **c**, VSMCs were incubated with or without pIFN- γ , PDGF-BB (10 ng ml⁻¹) or IL-1 β (1 ng ml⁻¹) and counted after 3 days. **d**, VSMCs were cultured in 0.5% serum in the presence or absence of pIFN- γ for 2 days, then incubated with or without PDGF-BB (10 ng ml⁻¹) and counted after 3 days. **e**, **f**, Saturation binding (**e**) and Scatchard analysis (**f**) of ¹²⁵I-PDGF-BB bound to VSMCs grown in 0.5% serum with or without pIFN- γ for 2 days. Data represent mean $\pm$ s.e.m. (cell counting, $n = 6$; radioligand binding, $n = 4$). Statistical significance was determined by one-way analysis of variance. * $P < 0.001$, ** $P < 0.01$.

of PDGF receptors. Immunohistochemical staining of cultured VSMCs revealed an upregulation of PDGF β -receptors by pIFN- γ treatment (data not shown). Binding studies using radiolabelled PDGF-BB established that pIFN- γ exposure increased the number of surface PDGF receptors on VSMCs without altering the ligand affinity (Fig. 5e, f), which explains, at least in part, the increase in cellular proliferation. We are currently investigating other possible interactions between IFN- γ and PDGF signalling molecules and target gene products. Overall, the *in vitro* data support the conclusion indicated by our *in vivo* analyses that the mitogenic effect of IFN- γ is mediated through PDGF.

Our results lead to two conclusions about the effects of IFN- γ on vascular tissue. First, arterial endothelial expression of MHC class I and II antigens is not constitutive, disappears *in vivo* in the absence of species-specific factors, and is sustained by exogenous IFN- γ . The pattern and degree of MHC antigen expression in vascular grafts from cytokine-treated animals seemed to be physiological, although the model does not necessarily reproduce the cellular exposure to endogenously produced IFN- γ . These observations may explain previous findings that capillary endothelial expression of MHC class II antigens in dogs is suppressed by cyclosporine²¹ and that arterial endothelial cells from mice deficient in IFN- γ or IFN- γ receptors

fail to express MHC class I molecules²². Endothelial MHC antigens are recognized by T cells, and their dependence on IFN- γ for basal expression indicates that IFN- γ has a physiological role in non-inflammatory states, for example, by conducting immune surveillance for microbial pathogens. A further implication is that xenografts, unlike allografts, may become progressively less immunogenic in a host of a different species.

Second, we found that IFN- γ acts on transplanted segments of artery to produce arteriosclerotic changes in the absence of detectable leukocytes by potentiating PDGF-induced mitogenesis. This was unexpected because most studies have shown that IFN- γ inhibits the growth and biosynthetic capacity of VSMCs^{9–15} and suppresses endothelial cell PDGF production²³, although there have been reports of IFN- γ promoting VSMC growth^{24,25} through up-regulation of PDGF β -receptors²⁴. In accord with our results on VSMCs, IFN- γ results in the proliferation of some other mesenchymal cell types^{26–29} by mechanisms including increased DNA synthesis in response to PDGF²⁹. There are several important differences between our animal model and previous studies that reported inhibition of VSMC growth^{9–18}, including the prolonged duration of cytokine treatment and the preservation of normal vascular architecture. In this regard, our experiments are more similar to reports of atherosclerosis or graft arteriosclerosis in which IFN- γ was found to be necessary for intimal expansion^{4–8}. These studies have generally interpreted the effect of IFN- γ as being immunomodulatory. Our finding that IFN- γ has an arteriosclerotic effect in the apparent absence of immunocytes indicates that these studies may need to be reinterpreted. The observation that IFN- γ increases the mitogenic activity of PDGF may also apply to other chronic inflammatory disease processes that are characterized by the presence of activated lymphocytes, growth factors and mesenchymal cell proliferation.

Methods

Artery transplantation

Size-matched segments of epicardial coronary arteries from Yorkshire pigs (30–60 kg), or of coronary, inferior epigastric and internal mammary arteries from adult humans were transplanted as interposition grafts into the infrarenal aorta of 'non-leaky' (serum IgG levels $< 1 \mu\text{g ml}^{-1}$) C.B-17 SCID/beige mice (Taconic, 6–8 weeks old) using an end-to-end microsurgical technique³⁰.

Recipient treatment

Mice were treated with recombinant pIFN- γ (VIDO), heat-inactivated pIFN- γ (at 90 °C for 15 min) or recombinant hIFN- γ (R&D) at 200 ng, subcutaneously, three times a week, starting one week post-operatively and continuing for up to four weeks. Cytokine administration was delayed for one week after transplantation to allow resolution of inflammation around the sutures and complete emigration of pig passenger leukocytes from the vascular graft. Pyrogen-free saline was administered to control mice. Some animals received BrdU (Sigma) injected subcutaneously (100 mg per kg) and intraperitoneally (30 mg per kg) at 18 hours, and intraperitoneally (30 mg per kg) at 12 hours, before being killed.

Histology and immunohistochemistry

Mice were perfused with 2% paraformaldehyde in phosphate-buffered saline (PBS) through the left ventricle, and the grafts were removed. Haematoxylin and eosin (H&E) and elastin Van Gieson (EVG) stains were performed using standard techniques. Alternatively, recipients were perfused with PBS alone. Immunohistochemical staining used the following primary antibodies: mouse anti-BrdU (Becton Dickinson); mouse anti-human smooth muscle α -actin and PCNA (Sigma); mouse anti-pig MHC class I antigen, MHC class II antigen (VMRD), PECAM-1, CD45 (Serotec) and PDGF β -receptor (Sigma); rat anti-mouse MHC class I antigen, PECAM-1 and CD45 (Pharmingen); goat anti-human PDGF-AB; rabbit anti-human or anti-pig TGF- β 1 or TGF- β 2; and rabbit anti-human TGF- β type-II receptor (R&D). The monoclonal antibodies against pig and mouse antigens displayed species-specific staining. The polyclonal antibodies to human antigens cross-reacted with the corresponding porcine molecules. Binding of secondary antibodies (Jackson ImmunoResearch) was detected with the peroxidase and 3-amino ethyl carbazole kits (Vector).

Morphometry

Mice were perfused at a pressure of 100 mm Hg with PBS, 0.1 mM adenosine and 0.3 mM papaverine (Sigma) for 5 min at 37 °C, followed by perfusion fixation with 4% paraformaldehyde for 3 min. The grafts were excised and postfixed for 24 hours at 4 °C. H&E and

EVG stains identified the vascular layers, and morphometric analysis was performed using video microscopy. Under the assumption that the vessel cross-sections were circular, the perimeters of the endothelium, internal (IEL) and external elastic laminae (EEL) were outlined and area measurements calculated using an image software program (Scion). The lumen area (within the endothelium), intima area (between the endothelium and IEL) and media area (between the IEL and EEL) of five serial cross-sections 150 μm apart were assessed and averaged for each vessel. The maximal intima thickness (from endothelium to IEL) was determined from longitudinal sections. Nuclei were counted on microscopic examination of haematoxylin-stained cross-sections.

Cell growth analysis

Pig VSMCs were isolated from explant outgrowths of collagenase-digested, minced epicardial coronary arteries after mechanical removal of the adventitia and endothelial cells. The cells were cultured in M199 supplemented with 10% fetal bovine serum (FBS), antibiotics and 3 mM L-glutamine (Life Technologies). Experiments were carried out between cell passages 2–4. The VSMCs were subcultured into 24-well plates (Becton Dickinson) at a subconfluent density of 1×10^4 cells cm^{-2} and initially incubated in 0.5% FBS for 48 hours to induce growth quiescence. Various concentrations of pIFN- γ , purified porcine PDGF-BB (R&D), TGF- β 1 (R&D) or recombinant IL-1 β (Endogen) were added in medium supplemented with 0.5% or 10% FBS. The media and additives were replaced every 2–3 days. Adherent VSMCs were collected after treatment with 0.05% trypsin and 0.5 mM EDTA, and counted using a haemocytometer. The DNA content of ethanol-fixed cells incubated with propidium iodide (Boehringer Mannheim) at 50 $\mu\text{g ml}^{-1}$ for 15 min was analysed using a FACScan (Becton Dickinson).

Radiolabelled PDGF binding

VSMCs were cultured in 0.5% serum with or without pIFN- γ for 48 hours and washed in PBS, 1 mM CaCl_2 and 0.1% bovine serum albumin (BSA). Various concentrations of recombinant human ^{125}I -PDGF-BB (Amersham) were added in 1 ml of DMEM, 25 mM HEPES, 0.25% BSA for 4 hours at 4 °C. After four further washes, the cells were lysed in 1 ml of 1% Triton X-100, 0.1% BSA, H_2O and transferred to a gamma counter. Nonspecific binding was calculated with excess unlabelled ligand, assuming a linear relation with ^{125}I -PDGF-BB concentrations. Saturation binding was determined and Scatchard analysis performed using a GraphPad program.

Data analysis

Experiments were done at least three times with reproducible results. Data are expressed as means $\pm$ s.e.m. Experimental groups of animals were compared using the Mann–Whitney U test. Cell culture experimental groups were compared by one-way analysis of variance. The significance of post-hoc comparisons was tested by the Scheffé method where appropriate.

Received 16 August; accepted 8 November 1999.

- Ross, R. Atherosclerosis—an inflammatory disease. *N. Engl. J. Med.* **340**, 115–126 (1999).
- Hansson, G. K., Jonasson, L., Seifert, P. S. & Stemme, S. Immune mechanisms in atherosclerosis. *Arteriosclerosis* **9**, 567–578 (1989).
- Libby, P., Salomon, R. N., Payne, D. D., Schoen, F. J. & Pober, J. S. Functions of vascular wall cells related to development of transplantation-associated coronary arteriosclerosis. *Transplant. Proc.* **21**, 3677–3684 (1989).
- Russell, P. S., Chase, C. M., Winn, H. J. & Colvin, R. B. Coronary atherosclerosis in transplanted mouse hearts. III. Effects of recipient treatment with a monoclonal antibody to interferon- γ . *Transplantation* **57**, 1367–1371 (1994).
- Gupta, S. et al. IFN- γ potentiates atherosclerosis in ApoE knock-out mice. *J. Clin. Invest.* **99**, 2752–2761 (1997).
- Nagano, H. et al. Interferon- γ deficiency prevents coronary arteriosclerosis but not myocardial rejection in transplanted mouse hearts. *J. Clin. Invest.* **100**, 550–557 (1997).
- Räisänen-Sokolowski, A., Glysing-Jensen, T., Koglin, J. & Russell, M. E. Reduced transplant arteriosclerosis in murine cardiac allografts placed in interferon- γ knockout recipients. *Am. J. Pathol.* **152**, 359–365 (1998).
- Nagano, H. et al. Coronary arteriosclerosis after T-cell-mediated injury in transplanted mouse hearts. Role of interferon- γ . *Am. J. Pathol.* **152**, 1187–1197 (1998).
- Hansson, G. K., Jonasson, L., Holm, J., Clowes, M. M. & Clowes, A. W. γ -Interferon regulates vascular smooth muscle proliferation and Ia antigen expression *in vivo* and *in vitro*. *Circ. Res.* **63**, 712–719 (1988).
- Warner, S. J. C., Friedman, G. B. & Libby, P. Immune interferon inhibits proliferation and induces 2'-5'-oligoadenylate synthetase gene expression in human vascular smooth muscle cells. *J. Clin. Invest.* **83**, 1174–1182 (1989).
- Hansson, G. K., Hellstrand, M., Rymo, L., Rubbia, L. & Gabbiani, G. Interferon γ inhibits both proliferation and expression of differentiation-specific α -smooth muscle actin in arterial smooth muscle cells. *J. Exp. Med.* **170**, 1595–1608 (1989).
- Shimokado, K. & Numano, F. Inhibition of human vascular smooth muscle cell proliferation by interferon gamma. *Ann. N. Y. Acad. Sci.* **598**, 544–545 (1990).
- Nunokawa, Y. & Tanaka, S. Interferon- γ inhibits proliferation of rat vascular smooth muscle cells by nitric oxide generation. *Biochem. Biophys. Res. Commun.* **188**, 409–415 (1992).
- Bennett, M. R., Evan, G. I. & Newby, A. C. Deregulated expression of the *c-myc* oncogene abolishes inhibition of proliferation of rat vascular smooth muscle cells by serum reduction, interferon- γ , heparin, and cyclic nucleotide analogues and induces apoptosis. *Circ. Res.* **74**, 525–536 (1994).
- Amento, E. P., Ehsani, N., Palmer, H. & Libby, P. Cytokines and growth factors positively and negatively regulate interstitial collagen gene expression in human vascular smooth muscle cells. *Arterioscler. Thromb.* **11**, 1223–1230 (1991).
- Hansson, G. K. & Holm, J. Interferon- γ inhibits arterial stenosis after injury. *Circulation* **84**, 1266–1272 (1991).

- Hansson, G. K. et al. T lymphocytes inhibit the vascular response to injury. *Proc. Natl. Acad. Sci. USA* **88**, 10530–10534 (1991).
- Castronuovo, J. J. et al. Cytokine therapy for arterial restenosis: inhibition of neointimal hyperplasia by gamma-interferon. *Cardiovasc. Surg.* **3**, 463–468 (1995).
- Mach, F., Schönbeck, U., Sukhova, G. K., Atkinson, E. & Libby, P. Reduction of atherosclerosis in mice by inhibition of CD40 signalling. *Nature* **394**, 200–203 (1998).
- Sultan, P. et al. Pig but not human interferon- γ initiates human cell-mediated rejection of pig tissue *in vivo*. *Proc. Natl. Acad. Sci. USA* **94**, 8767–8772 (1997).
- Groenewegen, G., Buurman, W. A. & van der Linden, C. J. Lymphokine dependence of *in vivo* expression of MHC class II antigens by endothelium. *Nature* **316**, 361–363 (1985).
- Goes, N., Urmson, J., Hobart, M. & Halloran, P. F. The unique role of interferon- γ in the regulation of MHC expression on arterial endothelium. *Transplantation* **62**, 1889–1894 (1996).
- Suzuki, H. et al. Interferon- γ modulates messenger RNA levels of *c-sis* (PDGF-B chain), PDGF-A chain, and IL-1 β genes in human vascular endothelial cells. *Am. J. Pathol.* **134**, 35–43 (1989).
- Yokota, T. et al. Mitogenic activity of interferon gamma on growth-arrested human vascular smooth muscle cells. *Arterioscler. Thromb.* **12**, 1393–1401 (1992).
- Mensink, A. et al. Modulation of intercellular communication between smooth muscle cells by growth factors and cytokines. *Eur. J. Pharmacol.* **310**, 73–81 (1996).
- Brinckerhoff, C. E. & Guyre, P. M. Increased proliferation of human synovial fibroblasts treated with recombinant immune interferon. *J. Immunol.* **134**, 3142–3146 (1985).
- Hunninghake, G. W., Hemken, C., Brady, M. & Monick, M. Immune interferon is a growth factor for human lung fibroblasts. *Am. Rev. Respir. Dis.* **134**, 1025–1028 (1986).
- Yong, V. W. et al. γ -Interferon promotes proliferation of adult human astrocytes *in vitro* and reactive gliosis in the adult mouse brain *in vivo*. *Proc. Natl. Acad. Sci. USA* **88**, 7016–7020 (1991).
- Marra, F., Choudhury, G. G. & Abboud, H. E. Interferon- γ -mediated activation of STAT1 α regulates growth factor-induced mitogenesis. *J. Clin. Invest.* **98**, 1218–1230 (1996).
- Lorber, M. I. et al. Human allogeneic vascular rejection after arterial transplantation and peripheral lymphoid reconstitution in severe combined immunodeficient mice. *Transplantation* **67**, 897–903 (1999).

Acknowledgements

This work was supported by grants from the American Heart Association (G.T.) and the NIH (J.S.P.). R.W.K. received a fellowship award for the Thoracic Surgery Foundation for Research and Education, and J.S.S. was supported by a career development award from the Dermatology Foundation.

Correspondence and requests for materials should be addressed to G.T. (e-mail: george.tellides@yale.edu).

Negative regulation of lymphocyte activation and autoimmunity by the molecular adaptor Cbl-b

Kurt Bachmaier*, Connie Krawczyk*, Ivona Kozieradzki*, Young-Yun Kong*, Takehiko Sasaki*, Antonio Oliveira-dos-Santos*, Sanjeev Mariathasan†, Dennis Bouchard*, Andrew Wakeham*, Annick Itie*, Jenny Le*, Pamela S. Ohashi†, Idiko Sarosi‡, Hiroshi Nishina*, Stan Lipkowitz§ & Josef M. Penninger*

*Amgen Institute, Ontario Cancer Institute, Departments of Medical Biophysics and Immunology, University of Toronto, Toronto, Ontario M5G 2C1, Canada

†Ontario Cancer Institute, Departments of Medical Biophysics and Immunology, University of Toronto, Toronto, Ontario M5G 2M9, Canada

‡Department of Pathology, Amgen Inc., Thousand Oaks, California 91320-1789, USA

§Genetics Department, Medicine Branch, National Cancer Institute, Bethesda Naval Hospital, Bethesda, Maryland 20889-5101, USA

The signalling thresholds of antigen receptors and co-stimulatory receptors determine immunity or tolerance to self molecules¹. Changes in co-stimulatory pathways can lead to enhanced activation of lymphocytes and autoimmunity, or the induction of clonal anergy². The molecular mechanisms that maintain immunotolerance *in vivo* and integrate co-stimulatory signals with antigen receptor signals in T and B lymphocytes are poorly understood. Members of the Cbl/Sli family of molecular adaptors function downstream from growth factor and antigen receptors^{3–5}. Here we show that gene-targeted mice lacking the adaptor Cbl-b develop spontaneous autoimmunity characterized by auto-antibody

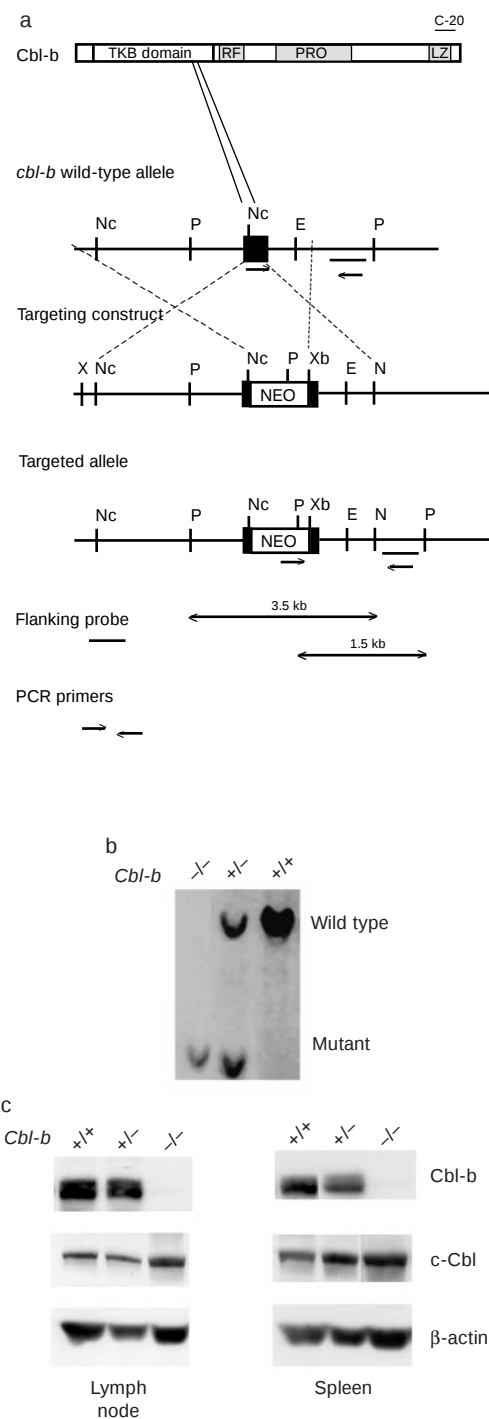


Figure 1 Gene targeting of *cbl-b*. **a**, Representation of the structural domains of Cbl-b and the genomic *cbl-b* sequences, and construction of the neomycin resistance (Neo) vector. The *cbl-b* exon is shown as a box. The *cbl-b* flanking probe used for Southern blotting, expected fragment sizes after digests of wild-type (3.5 kb) and mutant (1.5 kb) genomic DNA, and location of PCR primers are indicated. Nc, *Nco*I; P, *Pst*I; E, *Eco*RI; X, *Xba*I; Xb, *Xba*I; N, *Not*I. The deleted region spanning residues 300–340, located in the N-terminal portion of Cbl-b within the tyrosine kinase-binding (TKB) domain and 5' of the RING finger (RF) domain, is indicated. LZ, levline zipper; PRO, proline-rich domain. **b**, Genomic DNA was isolated from *cbl-b*^{+/+}, *cbl-b*^{+/-} and *cbl-b*^{-/-} mice, digested with *Pst*I and analysed by Southern blotting using the 3' flanking probe shown in Fig. 1a. Wild-type and mutant bands are indicated. **c**, Western blot analysis of Cbl-b protein expression in *cbl-b*^{+/+}, *cbl-b*^{+/-} and *cbl-b*^{-/-} lymph-node cells and splenocytes. Whole-cell lysates (20 μ g) were probed with an anti-Cbl-b antibody (C-20 shown in Fig. 1a). Cbl-b protein appears in two bands because of alternative splicing⁵. β -actin and c-Cbl are shown as controls.

production, infiltration of activated T and B lymphocytes into multiple organs, and parenchymal damage. Resting *cbl-b*^{-/-} lymphocytes hyperproliferate upon antigen receptor stimulation, and *cbl-b*^{-/-} T cells display specific hyperproduction of the T-cell growth factor interleukin-2, but not interferon- γ or tumour necrosis factor- α . Mutation of Cbl-b uncouples T-cell proliferation, interleukin-2 production and phosphorylation of the GDP/GTP exchange factor Vav1 from the requirement for CD28 co-stimulation. Cbl-b is thus a key regulator of activation thresholds in mature lymphocytes and immunological tolerance and autoimmunity.

As the role of Cbl-b in T-cell activation *in vitro* is controversial^{6–8}, and nothing is known about the *in vivo* function of Cbl-b, we generated mice that were deficient in the *cbl-b* gene (Fig. 1a). *cbl-b*^{-/-} mice (Fig. 1b) were fertile and born at mendelian frequency. Messenger RNA from *cbl-b* (data not shown) and Cbl-b protein were absent in *cbl-b*^{-/-} lymph-node cells and splenocytes (Fig. 1c). At 6 months of age, *cbl-b*^{-/-} mice exhibited a huge mass in the submandibular salivary glands (Fig. 2a,b) that resulted from an

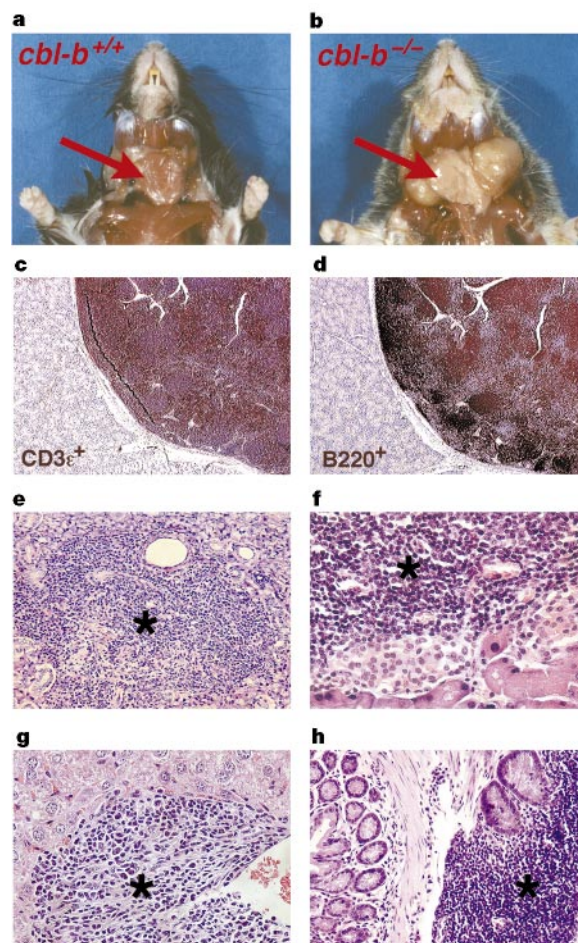


Figure 2 *cbl-b*^{-/-} mice develop spontaneous, generalized autoimmunity. **a, b**, Anterior neck regions of 14-month-old *cbl-b*^{+/+} and *cbl-b*^{-/-} mice. **a**, Normal anatomy of the submandibular salivary glands in wild-type mouse (arrow). **b**, A large mass in the place of the submandibular salivary glands is apparent in the *cbl-b*^{-/-} mouse (arrow). **c, d**, Histological cross-sections of lymphoid neo-organogenesis in the submandibular salivary gland of a *cbl-b*^{-/-} mouse. Immunostaining with anti-CD3 ϵ antibody for T cells (**c**) and with anti-B220 antibody for B cells (**d**). Similar lymphoid neo-organogenesis was observed in the lung and the pancreas of *cbl-b*^{-/-} mice. **e–h**, Mononuclear cellular infiltrate and tissue destruction (asterisk) in various organs of a 6-month-old *cbl-b*^{-/-} mice visualized by HE staining. **e**, Submandibular salivary gland. **f**, Exocrine and endocrine pancreas. **g**, Central vein region of the liver. **h**, Large intestine. Original magnifications $\times 4$ (**b, d**), $\times 20$ (**f, g**), $\times 40$ (**i, k**), $\times 30$ (**h, j**).

accumulation of T and B cells and lymphoid neo-organogenesis^{9,10} (Fig. 2c,d). Moreover, the salivary glands, exocrine and endocrine pancreas, liver, intestine, lung, kidney, heart, skeletal muscle, urinary bladder and connective tissues were infiltrated by mononuclear cells with accompanying parenchymal damage (Fig. 2e–h). Infiltration was variable in 3-month-old *cbl-b*^{-/-} mice but consistently present in all *cbl-b*^{-/-} mice (*n*=28) aged 6 months or older. In *cbl-b*^{+/-} littermates (*n*=22), minor infiltration could be detected in the submandibular salivary glands and kidneys, but other organs examined were free of infiltration. In wild-type

littermates (*n*=7), no infiltration could be detected in any organ. Infiltrates consisted of CD19⁺B220⁺ B lymphocytes and activated (CD69⁺CD25⁺) CD4⁺ and CD8⁺ T lymphocytes. T- and B-cell infiltrates were polyclonal, as judged by staining with various anti-T-cell receptor (TCR) V β and anti- κ and anti- λ antibodies, and thus did not represent clonal transformation. Moreover, older (>6 months) *cbl-b*^{-/-} mice developed progressive changes in facial features and enhanced osteoclastogenesis (see Supplementary Information).

Despite severe organ infiltration, spleens and lymph nodes were

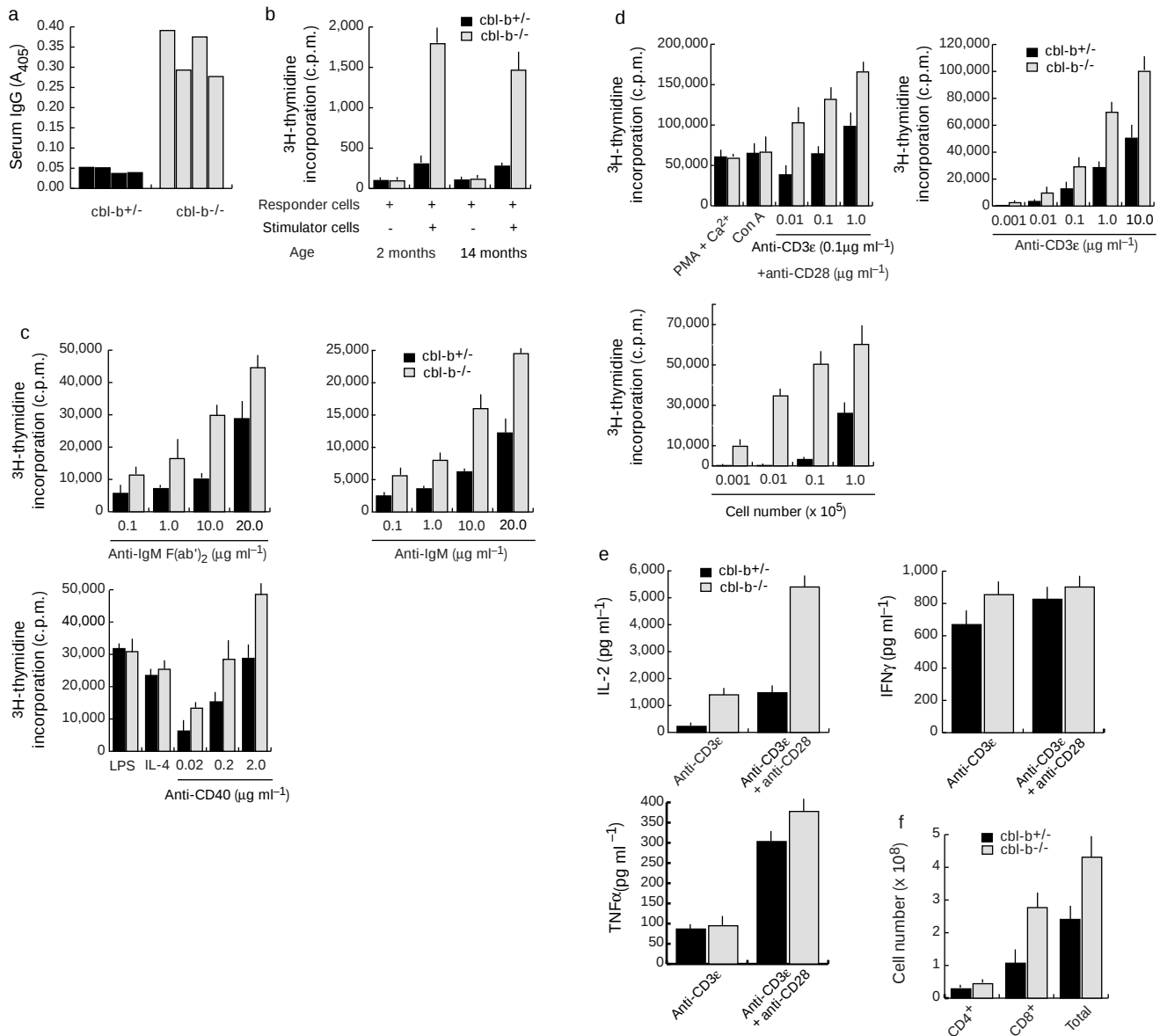


Figure 3 Self-reactive and hyperresponsive lymphocytes in *cbl-b*^{-/-} mice. **a**, Serum IgG auto-antibodies reactive to dsDNA determined by ELISA. Results from sera of individual 9-month-old *cbl-b*^{+/-} and *cbl-b*^{-/-} littermate mice are shown. **b**, Syngeneic MLRs using lymph-node T cells from a young (8-week-old) and an older (14-month-old) mouse as responder cells and γ -irradiated splenocytes as stimulator cells. Mean ³H-thymidine uptakes (counts per minute) $\pm$ s.d. are shown. **c**, Proliferation of splenic B cells (1×10^5) from young *cbl-b*^{+/-} and *cbl-b*^{-/-} mice (8-week-old) stimulated with the indicated doses of soluble anti-IgM F(ab')₂; soluble intact anti-IgM; soluble anti-CD40; LPS (2 μ g ml⁻¹); and recombinant interleukin (rIL-4) (400U ml⁻¹). Proliferation (³H-thymidine uptake $\pm$ s.d.) after 56 h stimulation. A representative result from a minimum of five experiments is shown. **d**, Proliferation of lymph-node T cells (1×10^5) from young (8-week-old) mice stimulated with soluble anti-CD3 ϵ (1 μ g ml⁻¹) plus the indicated doses of soluble anti-

CD28 antibodies, PMA (12.5ng ml⁻¹) plus Ca²⁺-ionophore (100ng ml⁻¹) or ConA (2.5 μ g ml⁻¹); proliferation of lymph-node T cells stimulated with different concentrations of soluble anti-CD3 ϵ antibody; and proliferation of lymph-node T cells plated at increasing cell density stimulated with soluble anti-CD3 ϵ (1 μ g ml⁻¹). **e**, IL-2, IFN γ and TNF α in supernatants of T cells from *cbl-b*^{+/-} and *cbl-b*^{-/-} mice stimulated for 24 h with soluble anti-CD3 ϵ antibody (1 μ g ml⁻¹) or soluble anti-CD3 ϵ (1 μ g ml⁻¹) plus anti-CD28 (0.1 μ g ml⁻¹). Cytokines were determined by ELISA. **f**, Expansion of splenocytes in young (8-week-old) *cbl-b*^{+/-} and *cbl-b*^{-/-} littermates (*n*=4 for each genotype) infected with LCMV (2 $\times 10^3$ p.f.u. intravenously). Splenocyte subpopulations were analysed by FACScan. Absolute numbers of splenocytes (total) and of CD4⁺ and CD8⁺ splenocytes are shown (mean $\pm$ s.d.).

only slightly enlarged (1.5–3 times) in older (>6 months) *cbl-b*^{-/-} mice. Lymph-node and splenic T cells of diseased *cbl-b*^{-/-} mice expressed high levels of CD69, CD25 and leukocyte function-associated antigen-1 (LFA-1) (data not shown), indicating that these cells had been activated *in vivo*. In addition, percentages and total numbers of Syndecan-1⁺B220^{low}CD23^{low} plasma cells¹¹ were increased in diseased *cbl-b*^{-/-} mice, and these mice exhibited anti-double stranded DNA (dsDNA) auto-antibodies (Fig. 3a). To determine whether spontaneously autoreactive T cells were present in *cbl-b*^{-/-} mice, we carried out syngeneic mixed lymphocyte reactions (MLRs). Lymph-node T cells from diseased *cbl-b*^{-/-} mice and naïve T cells from healthy (8-week-old) *cbl-b*^{-/-} mice responded with significantly stronger proliferation in syngeneic MLRs than T cells from heterozygous littermates (Fig. 3b). The generation of anti-dsDNA auto-antibodies, the proliferative response by T cells in syngeneic MLRs, the presence of inappropriately activated T and B cells in lymphatic organs, lymphoid neo-organogenesis, the infiltration of multiple organs with activated T and B cells, and tissue damage indicate that *cbl-b*^{-/-} mice develop a spontaneous, generalized autoimmune disorder.

Development of autoimmunity could be due to breakdown of self-tolerance in the thymus and/or changes in the activation thresholds of peripheral lymphocytes. In healthy *cbl-b*^{-/-} mice, differentiation of T and B cells, surface expression of TCR $\alpha\beta$, CD3, CD4, CD8, CD28, CD45, TCR $\beta\gamma$ subclasses, CD95, CD69, CD44, HSA, CD5, H2-K^b, IgM, IgD, CD23, CD19, CD80, CD86, CD40 and MHC class II (I-A^b), positive and negative selection of thymocytes expressing the MHC class I-restricted P14 and H-Y TCR transgenes, the numbers and ratios of T- and B-cell populations in lymph nodes and spleens, and susceptibility to apoptosis in response to anti-FAS (CD95), anti-CD3 ϵ plus anti-CD28, dexamethasone, sorbitol, or γ -irradiation were comparable to those of *cbl-b*^{+/+} mice (data not shown). Thus, in contrast to c-Cbl mutations^{12,13}, loss of Cbl-b has no notable effect on the development, selection and surface-receptor expression of T or B cells.

To determine the function of Cbl-b in peripheral T and B cells, we analysed their activation thresholds. *cbl-b*^{-/-} B cells displayed enhanced proliferative responses to anti-IgM, anti-IgM F(ab')₂ and anti-CD40 stimulation, but responded equally well to lipopolysaccharide (LPS) and IL-4 (Fig. 3c). Negative regulation of B-cell proliferation using intact anti-IgM, which interferes with B-cell receptor (BCR)-mediated activation by binding to the negative regulatory Fc γ IIb receptor, was not compromised (Fig. 3c). Lymph-node T cells from 8-week-old *cbl-b*^{+/+} and *cbl-b*^{-/-} T cells proliferated equally well in response to stimulation by phorbol myristate acetate (PMA)/Ca²⁺-ionophore or concanavalin A (ConA). However, *cbl-b*^{-/-} T cells were much more sensitive than control cells to anti-CD3 ϵ and anti-CD3 ϵ /anti-CD28 stimulation (Fig. 3d). *cbl-b*^{-/-} T cells responded at suboptimal concentrations of stimulating antibodies and at low cell densities. Intriguingly, stimulation with anti-CD3 ϵ alone triggered as much IL-2 in cultures of *cbl-b*^{-/-} T cells as stimulation of control T cells with anti-CD3 ϵ plus anti-CD28 (Fig. 3e). Interferon (IFN)- γ and tumour necrosis factor (TNF)- α levels did not differ significantly between *cbl-b*^{+/+} and *cbl-b*^{-/-} T cells (Fig. 3e). CD28 co-stimulation increased TCR/CD3-mediated proliferation and IL-2 production in *cbl-b*^{-/-} and *cbl-b*^{+/+} T cells (Fig. 3d,e). The kinetics and surface expression of CD25, CD69, CTLA4 and CD30, TCR internalization, receptor recycling and activation-induced apoptosis¹⁴ were comparable between *cbl-b*^{+/+}, *cbl-b*^{+/+} and *cbl-b*^{-/-} T cells in response to TCR/CD3 and CD28 (data not shown). After *in vivo* infection with lymphocytic choriomeningitis virus¹⁵ (LCMV), *cbl-b*^{-/-} mice exhibited significantly increased expansion of CD8⁺ cytotoxic T cells in spleens (Fig. 3f). Thus, Cbl-b adjusts the activation threshold for T and B cells in response to stimulation of antigen receptor.

Occupancy of the antigen receptor in the absence of co-stimulation

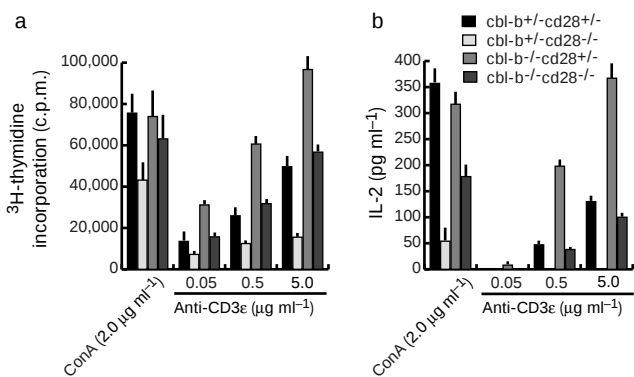


Figure 4 The *cbl-b* mutation uncouples T-cell proliferation and IL-2 production from CD28 co-stimulation. Proliferation (**a**) and IL-2 production (**b**) of lymph-node T cells (1×10^5) from 8-week-old *cbl-b*^{+/+}*cd28*^{-/-}, *cbl-b*^{+/+}*cd28*^{-/-}, *cbl-b*^{-/-}*cd28*^{+/+}, and *cbl-b*^{-/-}*cd28*^{-/-} littermate mice stimulated with ConA (2.5 µg ml⁻¹) or the indicated doses of soluble anti-CD3 ϵ . A representative result from three experiments is shown.

leads to anergy, which is defined by proliferative unresponsiveness and defective IL-2 production^{16–18}. Thus, potentially autoreactive T cells remain inactive because tissue cells do not provide the co-stimulation necessary for T-cell activation. TCR activation of naïve *cbl-b*^{-/-} T cells induced hyperproliferation and IL-2 production even in the absence of CD28 engagement. To test whether loss of Cbl-b can relieve a T cell from the requirement for CD28 co-stimulation, we introduced the *cbl-b* mutation into *cd28*^{-/-} mice; *cd28*^{-/-} T cells do not produce IL-2 and exhibit reduced proliferation following TCR/CD3 and ConA stimulation¹⁹. Development and peripheral populations of T and B cells were unaltered in *cd28*^{-/-}*cbl-b*^{-/-} mice. However, the *cbl-b* mutation restored T-cell proliferation and IL-2 production to wild-type levels in *cd28*^{-/-} mice upon TCR/CD3 or ConA stimulation (Fig. 4a,b). Thus, Cbl-b imposes a requirement for CD28 co-stimulation for proliferation and IL-2 production.

In primary murine T cells, Cbl-b associated with p85 α , phosphatidylinositol-3-OH kinase (PI(3)K), Slp76, ZAP70, p56^{lck}, phospholipase C (PLC)- γ -1 and the GDP/GTP exchange factor Vav1, but not with c-Cbl, Nck, PAK or the GTPases Rac1 and CDC42 (Fig. 5a). After TCR stimulation, the absence of Cbl-b had no significant effect on the kinetics or extent of overall tyrosine phosphorylation (data not shown), or on tyrosine phosphorylation of c-Cbl, ZAP70, PLC γ -1 and p56^{lck} (Fig. 5b). In *cbl-b*^{+/+} cells, significant tyrosine phosphorylation of Vav1 depends on TCR and CD28 co-engagement (Fig. 5b,c). Interestingly, Vav1 phosphorylation in *cbl-b*^{-/-} T cells achieved by CD3 stimulation alone was similar to that of *cbl-b*^{+/+} T cells stimulated by CD3 and CD28 (Fig. 5b,c). No significant effects on the kinetics or extents of SAPK/JNK, mitogen-activated kinase (MAPK)/extracellular signal-regulated kinase (ERK) or p38 kinase activation, or Ca²⁺ mobilization were observed in the absence of Cbl-b (data not shown). Thus, Cbl-b appears to be a negative regulator of Vav1 phosphorylation. The exact mechanism of Cbl-b-regulated Vav1 phosphorylation needs to be determined.

Whereas Chiang *et al.*²⁰ report increased autoimmunity in an experimental model, we observed the development of spontaneous autoimmunity in *cbl-b*-deficient mice. The onset and extent of spontaneous autoimmunity depends on age, genetic background and environmental factors. Our mice were kept under specific pathogen-free conditions and had been backcrossed for more than five generations into the C57BL/6 line, and, as noted above, histopathology is not apparent in young mice. Cbl-b emerges as a critical molecular adaptor that sets the signalling threshold and adjusts the requirement for co-stimulation in mature lymphocytes. We have shown that Cbl-b is involved in pivotal events of lymphocyte activation: (1) Cbl-b is a negative regulator of T- and B-cell activation; (2) Cbl-b regulates the triggering threshold of antigen

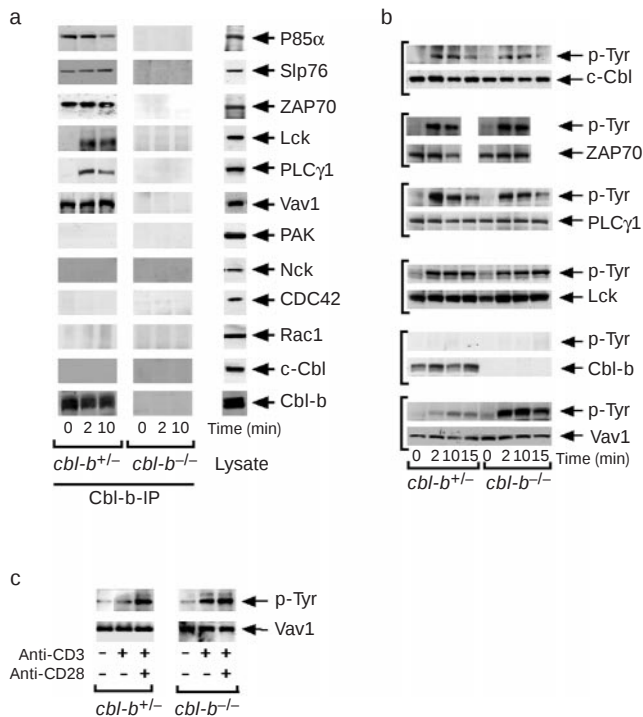


Figure 5 Cbl-b negatively regulates Vav1 tyrosine phosphorylation. **a**, Association between Cbl-b and selected signalling molecules in T cells. Wild-type T cells and, as a negative control, *cbl-b*^{-/-} T cells (1×10^7 per lane) were stimulated with anti-CD3ε ($1 \mu\text{g ml}^{-1}$) for the indicated durations. Cbl-b was immunoprecipitated and binding partners were identified using antibodies reactive to the indicated proteins. Right panels, base expression levels of the indicated proteins are shown for total lysates of non-activated lymph-node T cells (1×10^6 per lane). Base expression of the indicated proteins was comparable between *cbl-b*^{+/+}, *cbl-b*^{+/-} and *cbl-b*^{-/-} T cells (data not shown). Interactions between molecules were confirmed using three different anti-Cbl-b antibodies (Santa Cruz C-20, H121 and G-1), reverse immunoprecipitations, non-immune species-matched primary control antibodies and addition of specific peptides to deplete anti-Cbl-b antibodies. **b**, Enhanced tyrosine phosphorylation of Vav1 in *cbl-b*^{-/-} T cells. Purified lymph-node T cells (5×10^6 per lane) from 6-week-old *cbl-b*^{+/+} and *cbl-b*^{-/-} mice were stimulated with anti-CD3ε ($1 \mu\text{g ml}^{-1}$) for the indicated time periods, and indicated proteins were immunoprecipitated and immunoblotted for levels of expression and tyrosine phosphorylation (p-Tyr). Results were confirmed in IPs using anti-phosphotyrosine antibody. **c**, Enhanced tyrosine phosphorylation of Vav1 in response to anti-CD3ε stimulation in *cbl-b*^{-/-} T cells. Lymph-node T cells (5×10^6) from 6 wk old *cbl-b*^{+/+} and *cbl-b*^{-/-} mice were left untreated or stimulated with soluble anti-CD3ε ($1 \mu\text{g/ml}$) or soluble anti-CD3ε ($1 \mu\text{g ml}^{-1}$) plus anti-CD28 ($1 \mu\text{g ml}^{-1}$) for 15 min. Vav1 was immunoprecipitated and western blotted for levels of Vav1 and tyrosine phosphorylation (p-Tyr). A representative result from five experiments is shown.

receptors in T and B cells; (3) *cbl-b*^{-/-} T cells display hyperproduction of IL-2, but not TNFα or IFNγ; and (4) the *cbl-b* mutation uncouples T-cell proliferation, IL-2 production and Vav1 phosphorylation from the requirement for CD28 co-stimulation. All mutant mice spontaneously develop generalized autoimmunity that is characterized by T- and B-cell infiltration into vital organs. The Cbl-b-regulated signalling pathway in T and B cells may hold the key to our understanding of induction and/or maintenance of peripheral immunological tolerance and autoimmunity. □

Methods

Generation of *cbl-b* mutant mice

A 15-kilobase (kb) genomic *cbl-b* fragment was isolated from a 129/J mouse library and a targeting vector was constructed that contained a 526-base pair (bp) short arm and a 3.2-kb long arm of homology flanking a neomycin resistance cassette inserted in sense orientation to *cbl-b* transcription. The linearized construct was

electroporated into 1×10^7 E14K embryonic stem (ES) cells derived from 129/J mice. ES cell colonies resistant to G418 ($500 \mu\text{g ml}^{-1}$) were screened for homologous recombination by PCR using primers specific for *cbl-b* genomic sequences and the neomycin resistance cassette (sense primer: 5'-TTCCTCCCACTCATGATCATAG-3'; antisense primer: 5'-GGAAATATTAGTTCAACTGG-3'). Recombinant colonies were confirmed by Southern blotting of *Pst*I-digested genomic DNA hybridized to a 742-bp 3' flanking probe. Targeted cells were injected into blastocysts from C57BL/6 female mice. Chimeric male mice were crossed with C57BL/6 females to achieve germline transmission. After heterozygous matings, *cbl-b*^{-/-}, *cbl-b*^{+/-} and *cbl-b*^{+/+} mice were identified by PCR (wild-type primers: sense, 5'-CATCTCAGTGTGAAATTG-3'; antisense and mutant primers, as above) and confirmed by Southern blotting. For northern blotting, 10 μg of poly(A) RNA were electrophoresed on a 5.5% formaldehyde/1% agarose gel, transferred to a Hybond-N membrane (Amersham) and hybridized with probes specific for *cbl-b*, neomycin or β-actin. Cbl-b protein was detected by western blotting using immunopurified antibodies (Santa Cruz) directed against the following regions of Cbl-b: residues 668–692, C-20, polyclonal (goat origin); 600–721, H-121, polyclonal (rabbit); 29–483, G-1, monoclonal (mouse); 6–24, N-19, polyclonal (goat). The absence of Cbl-b protein in *cbl-b*^{-/-} cells was confirmed using those four different sets of antibodies that recognize the amino-terminal or carboxy-terminal regions of Cbl-b. Equivalent results and phenotypes were obtained for *cbl-b*^{-/-} mouse lines derived from two independently targeted ES cell clones. The *cbl-b* mutation was backcrossed onto a C57BL/6 background, and bred into LCMV p33 peptide-specific P14 and H-Y TCR transgenic mice^{21,22} and *cd28*^{-/-} mice¹⁹. Genotyping of mice for the P14 and H-Y TCR transgenes was carried out by staining peripheral blood lymphocytes with anti-TCRVβ8.1/8.2 and anti-Thy-1.2 antibodies. *cd28*^{-/-} mice were genotyped by PCR. Only littermate mice were used in all experiments. All mice were maintained at the animal facilities of the Ontario Cancer Institute under specific pathogen-free conditions according to institutional guidelines.

Histology and immunocytochemistry

Tissues were embedded in paraffin and sections were processed for staining with haematoxylin and eosin (HE). For immunocytochemical staining, frozen or deparaffinized sections ($0.4 \mu\text{m}$) were pre-incubated in a 3% H₂O₂ solution to block endogenous peroxidase activity and CAS Block (Zymed) to block nonspecific antibody-binding sites. Sections were incubated with primary antibodies for 1 h at room temperature. The following primary antibodies were used: anti-CD3ε (Dako); anti-CD4 (PharMingen), anti-CD8 (PharMingen); anti-B220 (PharMingen); anti-IgM (PharMingen); anti-IgD (PharMingen); anti-F4/80 (specific for macrophages; Serotec); anti-CD11b (Mac-1; PharMingen); anti-Igk and anti-Igλ (specific for immunoglobulin light chains). Biotinylated goat anti-rat immunoglobulin (BioGenex), biotinylated goat anti-rabbit immunoglobulin (Vector) and biotinylated goat anti-hamster immunoglobulin (Vector) were used as secondary reagents. Biotinylated antibodies were visualized using ABC reagent (Vector) and DAB enzyme substrates (Research Genetics). Sections were counterstained with haematoxylin. Non-immune species-matched immunoglobulins were used as negative controls for each primary antibody.

Immunocytometry

Single-cell suspensions of thymi, lymph nodes, bone marrow and spleens from *cbl-b*^{+/+}, *cbl-b*^{+/-} and *cbl-b*^{-/-} mice were stained with FITC-, PE- or biotin-conjugated antibodies reactive to Thy1.2, CD3ε, TCRαβ, TCRβ3, TCRβ6, TCRβ7, TCRβ8.1/8.2, TCRβ14, TCRα2, TCRγδ, CD4, CD8, CD25, CTLA4, CD28, CD95 (FAS), CD5, CD30, CD44, CD45, CD69, CD11a (LFA-1), CD80, CD86, CD40, MHC class II (I-A^b), CD23, B220, CD43, sIgM, sIgD, CD19 and Syndecan-1 (antibodies from PharMingen). For the analysis of thymocyte precursors, cell were stained with PE-conjugated anti-CD4, anti-CD8, anti-CD3ε, anti-B220, anti-CD11b, anti-Gr-1, anti-TCRαβ and anti-TCRγδ; FITC-conjugated anti-CD25; and biotin-conjugated anti-CD44. PE-negative precursor cells (triple-negative thymocytes) were analysed for expression of CD25 and CD44. Biotinylated antibodies were visualized using streptavidin-RED670 (Gibco BRL). All samples were analysed by flow cytometry using a FACScan (Becton Dickinson).

Lymphocyte proliferation and cytokine production

Lymphocytes were placed into 96-well plates in IMDM (10% FCS). T cells were stimulated with PMA (Sigma) plus the Ca²⁺ ionophore A23617, anti-CD3ε (clone 145-2C11, hamster IgG), anti-CD28 (clone 37.51, PharMingen) or ConA (Pharmacia). B cells were stimulated with anti-IgM (61-6800; Zymed), anti-IgM F(ab')₂ fragment (61-5900; Zymed), anti-CD40 (clone 145-2C11; PharMingen), murine rIL-4 (R&D Systems) or LPS (Sigma). Cells were stimulated in triplicate for different time periods and pulsed for the last 8 h with $1 \mu\text{Ci}$ per well ³H-thymidine (Amersham). ³H-thymidine incorporation was measured using a β-scintillation counter (Coulter). To determine cytokine production, cell culture supernatants were assayed in triplicate for the production of IL-2, IFNγ or TNFα using ELISA.

Syngeneic MLR, auto-antibody production and LCMV infection

For syngeneic MLR, 1×10^5 lymph-node responder cells and 4×10^5 γ-irradiated (20Gy) splenic stimulator cells from individual mice were cultured in 96-well flat-bottom plates in IMDM (10% FCS). Cells were cultured in triplicate for 5 days and ³H-thymidine incorporation was measured as above. Anti-dsDNA auto-antibodies were determined by ELISA using dsDNA-coated 96-well plates (The Binding Site). Bound serum IgG antibodies were detected using horseradish-peroxidase-conjugated antibody to mouse IgG. Substrate conversion was detected by measuring absorbance at 405 nm. For *in vivo* viral infections, mice were injected intravenously with LCMV (2×10^3 p.f.u.). Spleen cellularity and splenic populations were determined 8 days later using immunocytometry.

Signalling

T cells were stimulated with hamster anti-CD3 ϵ (clone 145-2C11, hamster IgG; PharMingen) or hamster anti-CD3 ϵ (clone 145-2C11) plus hamster anti-CD28 (clone 37.51; PharMingen) at 37°C for various times. Cells were lysed in ice-cold lysis buffer (10mM Tris-HCl pH 7.5, 1mM MgCl₂, 0.5% CHAPS, 1mM EDTA, 10% glycerol, 100 μ M Na₃VO₄, 50 mM NaF, 20 μ M leupeptin and 1mM benzamide). Whole-cell lysates and immunoprecipitates (IPs) were separated on SDS-polyacrylamide gels and proteins were transferred to nitrocellulose membranes and immunoblotted²³. Blots were visualized with enhanced chemiluminescence detection reagents (ECL; Amersham). Tyrosine phosphorylation was monitored in whole-cell lysates and IPs using an anti-phosphotyrosine monoclonal antibody (UBI) and antibodies reactive to p56^{lck}, ZAP70, Vav1, PLC γ -1, c-Cbl or Cbl-b (Santa Cruz, Transduction Laboratories and Zymed). Cbl-b IPs and whole-cell lysates were probed for the presence of different signalling molecules using antibodies specific for p85 α -PI3K, Slp76, ZAP70, p56^{lck}, PLC γ -1, Vav1, Pak, Nck, CDC42, Rac1 or c-Cbl (Santa Cruz, Transduction Laboratories and Zymed). Activation of SAPK/JNKs, p38 and MAPKs (ERKs) was detected using monoclonal antibodies specific for phospho-SAPK/JNK, phospho-p38 or phospho-MAPK (New England Biolabs). The levels of SAPK/JNK, p38 and MAPK protein were determined by immunoblotting. For *in vitro* kinase assays, cells were lysed in 0.5% NP-40 lysis buffer and MAPK/ERKs and SAPK/JNKs were immunoprecipitated for 1 h at 4°C using a mixture of polyclonal rabbit anti-ERK1/anti-ERK-2 IgG (Santa Cruz), or polyclonal rabbit anti-SAPK/JNK IgG reactive against all SAPK isoforms (Santa Cruz). SAPK/JNK and MAPK kinase assays were performed using GST-c-Jun or myelin basic protein (MBP) as *in vitro* substrates for SAPK/JNK and MAPK (ERK1/ERK2), respectively. For detection of cytosolic Ca²⁺, 1 $\times$ 10⁷ freshly isolated lymph-node T cells were loaded with 3mM INDO-1 (Molecular Probes) in IMDM, pH 7.4, for 1 h at 37°C. After INDO-1 loading, cells were incubated with anti-CD3 ϵ (clone 145-2C11, 10 μ M) for 20 min at 4°C. Cells were washed and anti-CD3 ϵ molecules were crosslinked using goat anti-hamster IgG (Jackson Laboratories) at various concentrations. The cytosolic Ca²⁺ flux of 1 $\times$ 10⁶ cells was recorded in real time using a FACS Vantage.

Received 10 September; accepted 12 November 1999.

- Healy, J. I. & Goodnow, C. C. Positive versus negative signalling by lymphocyte antigen receptors. *Annu. Rev. Immunol.* **16**, 645–670 (1998).
- Rudd, C. E. Adaptors and molecular scaffolds in immune cell signalling. *Cell* **96**, 5–8 (1999).
- Cory, G. O. C. *et al.* The protein product of the c-cbl proto-oncogene is phosphorylated after B cell receptor stimulation and binds the SH3 domain of Bruton's tyrosine kinase. *J. Exp. Med.* **182**, 611–615 (1995).
- Sawasdikosol, S. *et al.* Tyrosine-phosphorylated Cbl binds to Crk after T cell activation. *J. Immunol.* **157**, 110–116 (1996).
- Keane, M. M., RiveroLezcano, O. M., Mitchell, J. A., Robbins, K. C. & Lipkowitz, S. Cloning and characterization of cbl-b: a SH3 binding protein with homology to the c-cbl proto-oncogene. *Oncogene* **10**, 2367–2377 (1995).
- Bustelo, X. R., Crespo, P., LopezBarahona, M., Gutkind, J. S. & Barbacid, M. Cbl-b, a member of the Sli-1/c-Cbl protein family, inhibits Vav-mediated c-Jun N-terminal kinase activation. *Oncogene* **15**, 2511–2520 (1997).
- Elly, C. *et al.* Tyrosine phosphorylation and complex formation of Cbl-b upon T cell receptor stimulation. *Oncogene* **18**, 1147–1156 (1999).
- Zhang, Z., Elly, C., Qiu, L., Altman, A. & Liu, Y. -C. A direct interaction between the adaptor protein Cbl-b and the kinase Zap-70 induces a positive signal in T cells. *Curr. Biol.* **9**, 203–206 (1999).
- Ludewig, B., Odermatt, B., Landmann, S., Hengartner, H. & Zinkernagel, R. M. Dendritic cells induce autoimmune diabetes and maintain disease via *de novo* formation of local lymphoid tissue. *J. Exp. Med.* **188**, 1493–1501 (1998).
- Kratz, A., CamposNeto, A., Hanson, M. S. & Ruddle, N. H. Chronic inflammation caused by lymphotoxin is lymphoid neogenesis. *J. Exp. Med.* **183**, 1461–1472 (1996).
- Sanderson, R. D., Lalor, P. & Bernfield, M. B lymphocytes express and lose syndecan at specific stages of differentiation. *Cell Reg.* **1**, 27–35 (1989).
- Murphy, M. A. *et al.* Tissue hyperplasia and enhanced T-cell signalling via ZAP-70 in c-Cbl-deficient mice. *Mol. Cell Biol.* **18**, 4872–4882 (1998).
- Naramura, M., Kole, H. K., Hu, R. J. & Gu, H. Altered thymic positive selection and intracellular signals in Cbl-deficient mice. *Proc. Natl Acad. Sci. USA* **95**, 15547–15552 (1998).
- Radvanyi, L. G., Mills, G. B. & Miller, R. G. Religation of the T cell receptor after primary activation of mature T cells inhibits proliferation and induces apoptotic cell death. *J. Immunol.* **150**, 5704–5715 (1993).
- Bachmann, M. F., Speiser, D. E. & Ohashi, P. S. Functional maturation of an antiviral cytotoxic T-cell response. *J. Virol.* **71**, 5764–5768 (1997).
- Boussiotis, V. A., Gribben, J. G., Freeman, G. J. & Nadler, L. M. Blockade of the CD28 co-stimulatory pathway: a means to induce tolerance. *Curr. Opin. Immunol.* **6**, 797–807 (1994).
- Celis, E. & Saibara, T. Binding of T cell receptor to major histocompatibility complex class II-peptide complexes at the single-cell level results in the induction of antigen unresponsiveness (anergy). *Eur. J. Immunol.* **22**, 3127–3134 (1992).
- Harding, F. A., McArthur, J. G., Gross, J. A., Raulet, D. H. & Allison, J. P. CD28-mediated signalling co-stimulates murine T cells and prevents induction of anergy in T-cell clones. *Nature* **356**, 607–609 (1992).
- Shahinian, A. *et al.* Differential T cell costimulatory requirements in CD28-deficient mice. *Science* **261**, 609–612 (1993).
- Chiang, Y. J. *et al.* Cbl-b regulates the CD28 dependence of T-cell activation. *Nature* **403**, 216–220 (2000).
- Pircher, H. *et al.* T cell tolerance to Mls encoded antigens in T cell receptor V beta 8. 1 chain transgenic mice. *EMBO J.* **8**, 719–727 (1989).
- Teh, H. S. *et al.* Thymic major histocompatibility complex antigens and the alpha beta T-cell receptor determine the CD4/CD8 phenotype of T cells. *Nature* **335**, 229–233 (1988).
- Fischer, K. D. *et al.* Vav is a regulator of cytoskeletal reorganization mediated by the T-cell receptor. *Curr. Biol.* **8**, 554–562 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Saunders for scientific editing, and T.W. Mak, J. Sasaki, N. Joza, M. Crackover, E. Griffith, M. Cheng, Q. Liu and P. Liu for comments. K.B. is supported by grants from the Heart and Stroke Foundation of Canada and by Amgen. J.M.P. is supported by the Medical Research Council (MRC) and the National Cancer Institute (NCI) of Canada.

Correspondence and request for materials should be addressed to J.M.P. (e-mail: jpenning@amgen.com).

Cbl-b regulates the CD28 dependence of T-cell activation

Yungping J. Chiang^{*†}, Hemanta K. Kole^{*†}, Karen Brown[‡], Mayumi Naramura^{*}, Shigetomo Fukuhara[§], Ren-Ju Hu^{*}, Ihn Kyung Jang^{*}, J. Silvio Gutkind[§], Ethan Shevach[‡] & Hua Gu^{*}

^{*} Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12441 Parklawn Drive, Rockville, Maryland 20852, USA

[‡] National Institutes of Health, 9,000 Rockville Pike, Bethesda, Maryland 20892, USA

[§] Oral and Pharyngeal Cancer Branch, NIDCR, National Institutes of Health, Bethesda, Maryland 20892, USA

[†] These authors contributed equally to this work

Whereas co-stimulation of the T-cell antigen receptor (TCR) and CD28 triggers T-cell activation, stimulation of the TCR alone may result in an anergic state or T-cell deletion, both possible mechanisms of tolerance induction^{1,2}. Here we show that T cells that are deficient in the adaptor molecule Cbl-b (ref. 3) do not require CD28 engagement for interleukin-2 production, and that the Cbl-b-null mutation (Cbl-b^{-/-}) fully restores T-cell-dependent antibody responses in CD28^{-/-} mice. The main TCR signalling pathways, such as tyrosine kinases Zap-70 and Lck, Ras/mitogen-activated kinases, phospholipase C γ -1 and Ca²⁺ mobilization, were not affected in Cbl-b^{-/-} T cells. In contrast, the activation of Vav, a guanine nucleotide exchange factor for Rac1/Rho/CDC42, was significantly enhanced. Our findings indicate that Cbl-b may influence the CD28 dependence of T-cell activation by selectively suppressing TCR-mediated Vav activation. Mice deficient in Cbl-b are highly susceptible to experimental autoimmune encephalomyelitis, suggesting that the dysregulation of signalling pathways modulated by Cbl-b may also contribute to human autoimmune diseases such as multiple sclerosis.

The mammalian Cbl family consists of three members, c-Cbl, Cbl-b and Cbl-3, which contain multiple protein-protein interaction domains without any known enzymatic activity^{4,5}. It has been shown that c-Cbl is a negative regulator for the tyrosine kinases Syk and Zap70 (refs 6–9) and modulates the TCR signalling threshold for thymic positive selection⁹. Cbl-b may positively regulate Zap70 signalling in Jurkat cells¹⁰; however, its physiological role remains unclear.

We generated mice deficient in the *cbl-b* gene (Cbl-b^{-/-}) by inserting a neomycin resistance gene into the third exon of the *cbl-b* gene (Fig. 1a, b). The lack of *cbl-b* gene transcripts and residual or truncated Cbl-b protein in Cbl-b^{-/-} T cells was confirmed by polymerase chain reaction with reverse transcriptase (RT-PCR) and western blot analysis using an antibody specific for the epitope at the carboxy terminus of Cbl-b (Fig. 1c, d; and data not shown). Mutant mice were fertile and grossly normal. Inspection of the central and peripheral lymphoid organs revealed a normal representation of B and T lymphocytes, natural killer (NK) and myeloid

lineage cells. Notably, in contrast to thymocytes of *c-Cbl*-deficient mice^{8,9}, the *Cbl-b*^{-/-} thymocytes did not show an altered expression of TCR, CD69 or CD5 molecules. Further examination of T-cell development in the thymus was carried out by crossing the mutant mice to two TCR transgenic mouse strains, one specific for the male antigen (H-Y)¹¹ and the other for pigeon cytochrome *c* (5C.C7)¹². Thymic positive and negative selection of T cells in *Cbl-b*^{-/-},

H-Y TCR transgenic and *Cbl-b*^{-/-}, 5C.C7 TCR transgenic mice appeared to be normal (data not shown), indicating that *Cbl-b* is not essential for the selection process of T cells in thymus.

To determine whether *Cbl-b* deficiency influences T-lymphocyte function, we examined activation of peripheral T cells by measuring their proliferation response. Mutant T-cell proliferation was markedly higher than that of wild-type T cells when stimulated with anti-CD3 ϵ antibody alone (Fig. 2a). However, this difference completely disappeared when exogenous interleukin (IL)-2 was added to the culture during stimulation (Fig. 2b), indicating that the enhanced proliferation of the *Cbl-b*^{-/-} T cells may have been influenced by IL-2 secreted from these cells. Indeed, when stimulated with anti-CD3 ϵ antibody alone, the mutant T cells produced a significant amount of IL-2, whereas the wild-type T cells produced minimal IL-2 under the same stimulation conditions (Fig. 2c). The enhanced IL-2 production by *Cbl-b*^{-/-} T cells was unlikely to have been due to the existing activated T cells, as the percentages of T cells bearing the activation markers, such as CD69, CD25 and CD62L^{low}, were almost the same in the wild-type and the mutant mice (data not shown). The CTLA-4 inhibitory signalling pathway appeared to be normal in the mutant cells because crosslinking the CTLA-4 during the culture inhibited IL-2 production (data not shown). To confirm further that TCR stimulation alone can efficiently induce IL-2 production in the absence of *Cbl-b*, we examined IL-2 production in T cells from *Cbl-b*^{-/-}CD28^{-/-} mice. In contrast to the CD28^{-/-} T cells, which showed a complete defect of IL-2 secretion, the *Cbl-b*^{-/-}CD28^{-/-} T cells produced the same amount of IL-2 as that produced by the *Cbl-b*^{-/-} cells after anti-CD3 ϵ stimulation (Fig. 2d). Together, these results indicate that the requirement of CD28 engagement for IL-2 production is diminished in the absence of *Cbl-b* *in vitro*.

To determine whether *Cbl-b*^{-/-} T cells can bypass CD28 regulation *in vivo*, we analysed the anti-NP (4-hydroxy-3-nitrophenylacetyl) hapten antibody response that is dependent on helper-T cells in *Cbl-b*^{-/-}CD28^{-/-} mice (Fig. 2e). As has been previously observed¹³, anti-NP IgG1 and IgG2b antibodies were severely reduced in

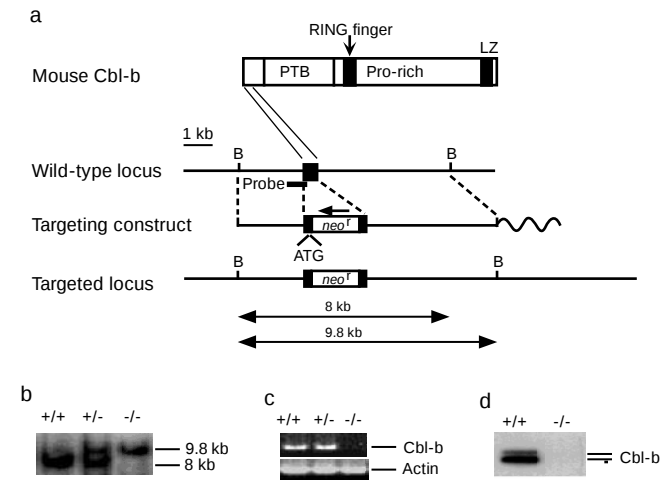


Figure 1 Generation of *Cbl-b*^{-/-} mice by gene targeting. **a**, Gene targeting strategy and the restriction map of the *cbl-b* gene. The solid box represents the third exon of the *cbl-b* gene with the translation initiation site. The neomycin resistance gene is inserted immediately after the translation initiation codon ATG. B, *Bam*HI. PTB, phosphotyrosine binding domain; LZ, leucine zipper domain. **b**, Southern blot analysis of mouse tail DNA. The 8-kb and 9.8-kb bands represent the germline and targeted allele, respectively. **c**, RT-PCR analysis of *cbl-b* gene expression. **d**, Western blot analysis of the *Cbl-b* protein. Anti-*Cbl-b* antibody is specific for the C-terminal portion of the *Cbl-b* protein.

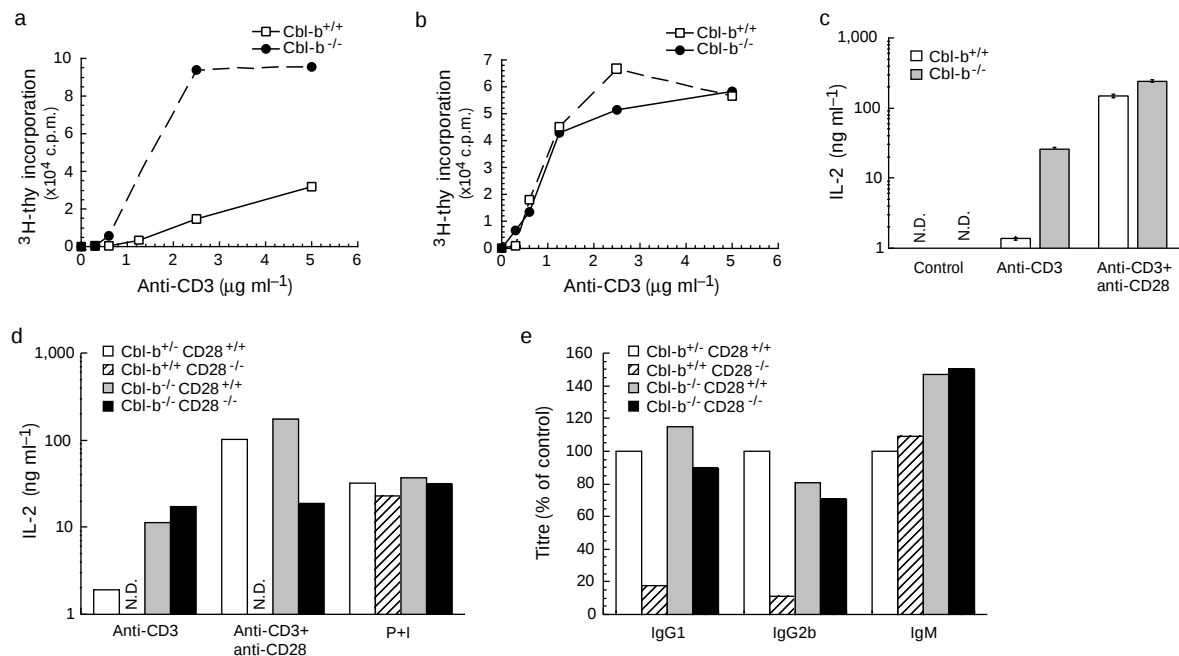


Figure 2 Comparison of T-cell proliferation and IL-2 secretion between *Cbl-b*^{-/-} and wild-type T cells. **a**, **b**, T-cell proliferation in the absence (**a**) or presence (**b**) of exogenous IL-2 (10 units ml⁻¹). The results represent the mean value of triplicate samples. thy, thymidine. **c**, IL-2 secretion by CD4⁺ T cells. The cells were cultured at concentration of 10⁶ cells ml⁻¹. The results represent the mean value of IL-2 concentration from triplicate samples. N.D., not detectable (<20 pg ml⁻¹). **d**, IL-2 production in wild-type, *Cbl-b*^{-/-}, CD28^{-/-} and *Cbl-b*^{-/-}CD28^{-/-} double mutant T cells. Purified T cells were stimulated with either plate-bound

anti-CD3 ϵ , anti-CD3 ϵ /CD28 or PMA/ionomycin (P+I) as indicated. The results represent the mean values of triplicate samples. **e**, Anti-NP specific antibody responses. The concentrations of anti-NP specific IgG1, IgG2b and IgM antibody were plotted as percentages of the control (*Cbl-b*^{+/+}, CD28^{+/+}) antibody concentration of the corresponding immunoglobulin isotypes. Results represent the mean values of antibody titres of 3–8 mice in each group.

Cbl-b^{+/+}CD28^{-/-} mice; however, the Cbl-b^{-/-}CD28^{-/-} mice mounted normal immunoglobulin (Ig)-G1 and IgG2b antibody responses. The IgM isotype anti-NP antibody response, which is not dependent on T-helper cells, did not change significantly in the double mutant mice, indicating that the increased IgG antibody responses in these mice were not caused by generally enhanced B-cell activation. Thus, activation of Cbl-b^{-/-} helper-T cells does not depend on the engagement of CD28 molecules *in vivo*.

Interleukin-2 expression and T-cell proliferation require activation of TCR proximal signalling components initiated by tyrosine kinases¹⁴, and the strong triggering of TCR alone can substitute CD28 signals and induce T-cell proliferation¹⁵. We therefore decided to examine whether signals transduced by the TCR were enhanced in the mutant cells. The tyrosine phosphorylations of CD3 ζ (p21 and p23), Zap70 and Lck tyrosine kinases were not affected in Cbl-b^{-/-} T cells (Fig. 3a; and data not shown). This result is inconsistent with a previous report that overexpression of Cbl-b in Jurkat cells enhances tyrosine phosphorylation of Zap-70 (ref. 10), and might reflect the difference between gene knockout and overexpression experiments. After anti-CD3 ϵ or anti-CD3 ϵ /CD4 antibody stimulation, the wild-type and mutant T cells displayed a similar pattern of tyrosine phosphorylation of total cellular proteins, the same degree of tyrosine phosphorylation on phospholipase C (PLC)- γ 1 and a similar level of Ca²⁺ mobilization (Fig. 3a, b; and see Supplementary Information). Furthermore, the wild-type and mutant T cells activated an equal amount of mitogen-activated protein kinases (MAPKs) extracellular signal-regulated kinase (ERK)1/ERK2 after TCR stimulation, indicating that the Ras/MAPK signalling pathway is not affected in the absence of Cbl-b (Fig. 3a). On the basis of these results, we conclude that Cbl-b deficiency does not seem to affect the above signalling pathways controlled by TCR.

It has been proposed that the Vav signalling pathway regulated by TCR or TCR/CD28 stimulation¹⁶. We found that Cbl-b was constantly associated with Vav in T cells. In the presence of Cbl-b, the TCR signalling alone had a minor role in eliciting tyrosine phosphorylation of Vav, whereas co-stimulation of the TCR/CD28 was

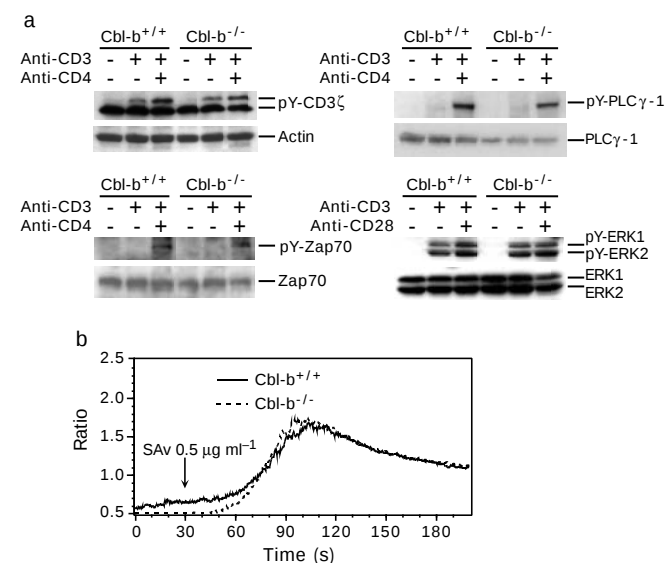


Figure 3 Biochemical analyses of signal transduction pathways in T cells from wild-type and Cbl-b^{-/-} mice. **a**, Western blot analysis of tyrosine phosphorylation (pY) on CD3 ζ , Zap70 and PLC γ -1, and active form of ERK1/2. T cells were purified from spleens and lymph nodes of wild-type (Cbl-b^{+/+}) and mutant (Cbl-b^{-/-}) mice, and stimulated with the indicated antibodies at 37 °C for 2 minutes. **b**, Ca²⁺ mobilization of wild-type and Cbl-b^{-/-} T cells after anti-CD3 ϵ antibody stimulation. Splenic T cells were pre-coated with biotinylated anti-CD3 ϵ antibody and then cross-linked with streptavidin (SAV). Data are presented as the median ratio of Ca²⁺-bound to Ca²⁺-free Indo-1 fluorescence as measured by flow cytometry.

necessary to induce maximal Vav phosphorylation (Fig. 4a). In contrast, stimulation with anti-CD3 ϵ or anti-CD3 ϵ /CD28 antibodies caused much stronger tyrosine phosphorylation of Vav in the Cbl-b^{-/-} T cells than in wild-type T cells (Fig. 4a). Cbl-b^{-/-} T cells already have a significant amount of tyrosine phosphorylated Vav even without stimulation, suggesting that Cbl-b may also restrain TCR-independent Vav activation in wild-type T cells. To determine further whether the enhanced tyrosine phosphorylation of Vav reflects the level of its guanine nucleotide exchange factor (GEF) activity, we examined Vav-mediated GDP/GTP exchange activity for Rac1 in the wild-type and mutant T cells. Consistent with the tyrosine phosphorylation data, stimulation of the wild-type cells through the TCR alone induced only minimal Vav GEF activity, whereas anti-CD3 ϵ /CD28 co-stimulation markedly enhanced GEF activity (Fig. 4b). In contrast, in the Cbl-b^{-/-} T cells, stimulation with anti-CD3 ϵ or anti-CD3 ϵ /CD28 antibody elicited much higher Vav GEF activity than that in the wild-type cells under the same stimulation conditions (Fig. 4b). Therefore, these results indicate that in T cells Cbl-b may inhibit TCR- and other receptor-mediated Vav activation, and that CD28 co-stimulation is necessary to overcome this inhibitory effect.

Signals delivered by CD28 have been proposed to be essential in T-cell-mediated immune self/non-self discrimination^{1,2}. CD28-independent activation of T cells in Cbl-b^{-/-} mice raises the possibility that these mice might have an increased susceptibility to autoimmune diseases. As the screening of Cbl-b^{-/-} mouse sera for the presence of auto-antibodies against double-stranded DNA did not reveal a meaningful difference between the wild-type and Cbl-b^{-/-} mice under 8 months of age (data not shown), we decided to examine this possibility using an experimental autoimmune disease

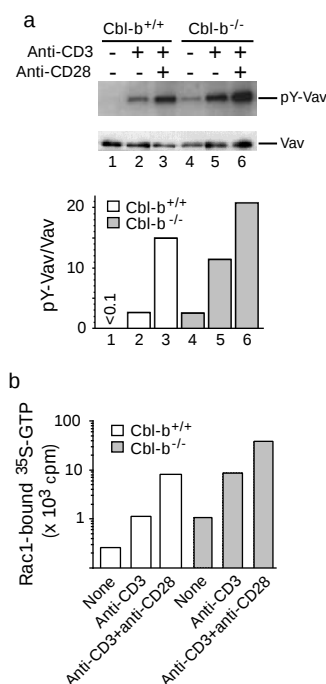


Figure 4 Analysis of Vav activation in Cbl-b^{-/-} T cells. **a**, Tyrosine phosphorylation of Vav in wild-type and Cbl-b^{-/-} T cells. Total Vav protein was immunoprecipitated from the cell lysates, transferred to PVDF membrane after SDS-PAGE and probed with anti-phosphotyrosine antibody (4G10) and a monoclonal anti-Vav antibody, respectively. Tyrosine-phosphorylated Vav and total Vav in each sample were determined by densitometric quantification. The bars represent the ratio of the intensity of tyrosine phosphorylated Vav compared with total Vav in the corresponding lanes. **b**, Vav-mediated Rac1-bound GDP/GTP exchange assay. Cells were stimulated as indicated. Total Vav protein immunoprecipitated from the cell lysate (2 × 10⁷ cells per sample) was used in the assay. The representative results of three independent experiments are shown.

Table 1 Clinical observation of immunized *cbl-b*^{-/-} and control mice

	Frequency of EAE	Clinical scores (no. of mice)			
		0	1	2	3
Control*	1 out of 13	12	1	0	0
<i>cbl-b</i> ^{-/-}	10 out of 13	3	5	1	4

Disease severity was scored on scale of 0 to 5: 0, no illness; 1, limp tail; 2, limp tail and hindlimb weakness; 3, hindlimb paralysis; 4, forelimb and hindlimb paralysis; 5, moribund.

* Control mice included 4 *cbl-b*^{+/+} and 9 *cbl-b*^{-/-} littermates.

model. Mouse experimental autoimmune encephalomyelitis (EAE) is an antigen-induced autoimmune disease mediated by autoreactive CD4⁺ T cells, and is considered to be a model of a human demyelinating disease, multiple sclerosis¹⁷. The mutant mice were highly susceptible to EAE, as ~80% (10 out of 13) of the *Cbl-b*^{-/-} mice became clinically ill after immunization with myelin basic protein. In contrast, only one sick mouse was found among a total of thirteen wild-type or heterozygous control littermates (Table 1).

Although our mice did not show significant signs of autoimmune disease, the authors of ref. 18 observed a high incidence of spontaneous autoimmune disease in their *Cbl-b*-deficient mice. It remains to be determined whether differences in the genetic backgrounds of the mice or their environments are responsible for this discrepancy. Nevertheless, the data from both groups show that deficiency of *Cbl-b* leads to impaired CD28 surveillance on T-cell activation and this is, at least in part, responsible for the enhanced susceptibility of the mutant animals to the autoimmune diseases. How does *Cbl-b* influence the CD28 dependence of IL-2 production during T-cell activation? We found that *Cbl-b* does not significantly affect the main TCR signalling pathways, including Zap70 and Lck tyrosine kinases, Ras/MAP kinases, PLCγ-1 and Ca²⁺ mobilization. Loss of *Cbl-b* correlates with Vav activation. It has been shown that the Vav signalling pathway plays a pivotal role in IL-2 expression^{19–23}. Over-expression of Vav in Jurkat T cells may enhance IL-2 expression²⁴. Thus, it is possible that *Cbl-b* influences CD28 dependence of T-cell activation by selectively restraining TCR mediated Vav activation in T cells.

CD28 dependence of T-cell activation has been proposed to play a central role in peripheral T-cell tolerance induction^{1,2}. Therefore, *Cbl-b*^{-/-} mice provide us with a unique model to dissect further the cellular and molecular mechanisms underlying this process. It will be interesting to examine whether the dysregulation of the *Cbl-b* pathway also contributes to other autoimmune diseases in humans. □

Methods

Generation of the *Cbl-b*^{-/-} mice

The genomic DNA of *cbl-b* gene was isolated from a 129 mouse genomic DNA library (Stratagene). An 8.0-kilobase (kb) *Bam*HI fragment containing the third exon of the *cbl-b* gene was used to make the targeting construct. The neomycin resistance gene was inserted into the third exon, immediately after the translation-initiation codon ATG, in the antisense orientation. The targeting construct was transfected into embryonic stem (ES) cells (E14–1) using a standard protocol, and homologous recombinants were identified by Southern hybridization. Two *cbl-b* gene targeted ES cell clones were identified and used to generate chimeric mice. Both lines were used as the founders to generate the homozygous *Cbl-b*^{-/-} mice. All the mice used in our experiments have a mixed genetic background between 129 and C57BL/6. To confirm the inactivation of the *cbl-b* gene in the homozygous mutant mice, total RNA was isolated from splenic cells and reverse transcribed with oligo-dT primers. The *cbl-b* and actin cDNA were amplified by PCR using the following primers: 5'-TTCCAGATGGCAACTCAATG-3' and 5'-TACATCTCTCCTTGCCTTCTTA-3' (primer pair for *cbl-b* cDNA); 5'-ATGCCAACACAGTGTCTCTGGTGG-3' and 5'-CTGATCCACTGTCTGGAAGGTG-3' (for actin). For western blot analysis, whole splenic cell lysate was blotted to a membrane, and the presence of *Cbl-b* protein was detected with polyclonal anti-*Cbl-b* antibody (C-20, Santa Cruz Biotechnologies).

Antibodies and reagents

Anti-CD3ε (2C11) and anti-CD28 (37.51) were purchased from Pharmingen. Antibodies against Zap70 (LR) were from Santa Cruz Biotechnologies. Anti-phosphotyrosine (4G10), rabbit polyclonal and mouse monoclonal antibodies against Vav, rabbit anti-MAPK antibody and monoclonal anti-PLCγ-1 antibody were from Upstate Biotechnology. Anti-active MAPK antibody was from Promega Corp.

T-cell proliferation and IL-2 assays

T cells and CD4⁺ T cells were purified from spleens or lymph nodes of 6–8-week-old mice, using T-cell and CD4⁺ T-cell separation columns (Accurate) according to the manufacturer's protocol. Flow cytometric analysis confirmed that the purity of the separated cells was consistently above 90%. CD4⁺, CD62L^{high} and CD62L^{low} T cells were purified on a FACS cell sorter, and the purity was greater than 98%. For the proliferation assay, 100 μl of purified CD4⁺ T cells (1 × 10⁶ cells ml⁻¹) were cultured in a 96-well (U-bottom) plate at 37 °C for 48 h. The plate was pre-coated with various concentrations of monoclonal anti-mouse CD3ε antibody. ³H-thymidine (1 μCi well⁻¹) was added at 48 h after stimulation and its incorporation was measured 16 h later. For IL-2 assay, 10⁵ cells (100 μl) were cultured in a 96-well plate pre-coated with 10 μg ml⁻¹ anti-CD3ε antibody, in the presence or absence of 1 μg ml⁻¹ soluble anti-CD28 antibody. The supernatant was collected 48 h later, and IL-2 in the culture media was measured using antibodies from Pharmingen. The results were obtained from triplicate samples.

Helper-T-cell-dependent antibody response

Six-to-eight-week-old mice were immunized by intraperitoneal injection of 50 μg NP-KLH (Solid Phase Sciences) precipitated in 100 μl of the adjuvant Inject Alum (1:4 dilution, Pierce). Mice were bled at 14 days after immunization. The anti-NP antibody responses were determined by enzyme-linked immunosorbent assay (ELISA) using plates coated with NP-bovine serum albumin. Bound antibodies of the IgG1, IgG2b and IgM isotypes were detected by alkaline-phosphatase-conjugated goat anti-mouse IgG1, IgG2b and IgM antibodies (Southern Biotechnological Associates), respectively. The antibody titres are presented as the percentages of those observed in normal mice.

Induction and evaluation of EAE.

Eight-to-ten-week-old mice were immunized with murine myelin basic protein as described²⁵. The clinical scores of the illness were evaluated according to Zamvil and Steinman¹⁷.

Western blotting, Ca²⁺ mobilization and Vav GDP/GTP exchange activity assays

T cells were isolated from lymph nodes and spleens using nylon-wool or magnetic columns and were of 85–90% purity as determined by flow cytometry. Antibody stimulation, immunoprecipitation and western blotting were carried out essentially as described²⁶. For the Vav-mediated GDP/GTP exchange assay, 2 × 10⁶ purified T cells were stimulated with 5–10 μg ml⁻¹ biotinylated anti-CD3ε or anti-CD3ε and anti-CD28 antibodies, and were crosslinked with 5–10 μg ml⁻¹ streptavidin, at 37 °C for 2 min. Total Vav protein was immunoprecipitated and subjected to the GEF activity assay using GST-Rac1 recombinant protein, as described²⁶. The amount of Vav used in each sample was quantified by western blot analysis. For Ca²⁺ mobilization analysis, purified T cells were loaded with Indo-1 AM (Molecular Probes) and stimulated with biotinylated anti-CD3ε antibody and streptavidin. Ca²⁺ mobilization was measured by flow cytometry.

Received 31 August; accepted 4 November 1999.

- Janeway, C. A. Jr & Bottomly, K. Signals and signs for lymphocyte responses. *Cell* **76**, 275–285 (1994).
- Schwartz, R. H. Models of T-cell anergy: is there a common molecular mechanism? *J. Exp. Med.* **184**, 1–8 (1996).
- Keane, M. M., Revero-Lezcano, O. M., Mitchell, J. A., Robbins, K. C. & Lipkowitz, S. Cloning and characterization of *cbl-b*: a SH3 binding protein with homology to the *c-cbl* proto-oncogene. *Oncogene* **10**, 2367–2377 (1995).
- Miyake, S. et al. The *Cbl* proto-oncogene product: from an enigmatic oncogene to center stage of signal transduction. *Crit. Rev. Oncol.* **8**, 189–219 (1997).
- Keane, M. M. et al. *Cbl-3*: A new mammalian *cbl* family protein. *Oncogene* **18**, 3365–3375 (1999).
- Ota, Y. & Samelson, L. E. The product of the proto-oncogene *c-cbl*: a negative regulator of the Syk tyrosine kinase. *Science* **276**, 418–420 (1997).
- Lupher, M. L. Jr, Songyang, Z., Shoelson, S. E., Cantley, L. C. & Band, H. The *Cbl* phosphotyrosine-binding domain selects a D(N/D)XPY motif and binds to the Tyr(292) negative regulatory phosphorylation site of Zap70. *J. Biol. Chem.* **272**, 33140–33144 (1997).
- Murphy, M. A. et al. Tissue hyperplasia and enhanced T-cell signaling via Zap70 in *c-Cbl*-deficient mice. *Mol. Cell. Biol.* **18**, 4872–4882 (1998).
- Naramura, M., Kole, H. K., Hu, R. J. & Gu, H. Altered thymic positive selection and intracellular signals in *Cbl*-deficient mice. *Proc. Natl Acad. Sci. USA* **95**, 15547–15552 (1998).
- Zhang, Z., Elly, C., Qiu, L., Altman, A. & Liu, Y. C. A direct interaction between the adaptor protein *Cbl-b* and the kinase Zap-70 induces a positive signal in T cells. *Curr. Biol.* **9**, 203–206 (1999).
- Kisielow, P., Bluthmann, H., Staerz, U. D., Steinmetz, M. & von Boehmer, H. Tolerance in T-cell-receptor transgenic mice involves deletion of nonmature CD4⁺CD8⁺ thymocytes. *Nature* **333**, 742–746 (1988).
- Seder, R. A., Paul, W. E., Davis, M. M. & Fazekas de St Grath, B. The presence of interleukin 4 during *in vitro* priming determines the lymphokine-producing potential of CD4⁺ T cells from T cell receptor transgenic mice. *J. Exp. Med.* **176**, 1091–1098 (1992).
- Shahinian, A. et al. Differential T cell costimulatory requirements in CD28-deficient mice. *Science* **261**, 609–612 (1993).
- Weiss, A. & Littman, D. R. Signal transduction by lymphocyte antigen receptors. *Cell* **76**, 263–274 (1994).
- Viola, A., Schroeder, S., Sakakibara, Y. & Lanzavecchia, A. T lymphocyte costimulation mediated by reorganization of membrane microdomains. *Science* **283**, 680–682 (1999).
- Cantrell, D. Lymphocyte signaling: a coordinating role for Vav. *Curr. Biol.* **8**, R535–R538 (1998).
- Zamvil, S. S. & Steinman, L. The T lymphocyte in experimental allergic encephalomyelitis. *Annu. Rev. Immunol.* **8**, 579–621 (1990).
- Bachmaier, K. et al. Negative regulation of lymphocyte activation and autoimmunity by the molecular adaptor *Cbl-b*. *Nature* **403**, 211–216 (2000).

19. Fischer K-D. *et al.* Vav is a regulator of cytoskeletal reorganization mediated by the T-cell receptor. *Curr. Biol.* **8**, 554–562 (1998).
20. Holsinger L. J. *et al.* Defects in actin-cap formation in Vav-deficient mice implicate an actin requirement for lymphocyte signal transduction. *Curr. Biol.* **8**, 563–572 (1998).
21. Zhang, A., Alt, F. W., Davidson, L., Orkin, S. H. & Swat, W. Defective signaling through T- and B-cell antigen receptors in lymphoid cells lacking the vav proto-oncogene. *Nature* **374**, 470–473 (1995).
22. Tarakhovskiy, A. *et al.* Defective antigen receptor-mediated proliferation of B and T cells in the absence of Vav. *Nature* **374**, 467–470 (1995).
23. Fischer, K. D. *et al.* Defective T-cell receptor signalling and positive selection of vav-deficient CD4⁺CD8⁺ thymocytes. *Nature* **374**, 474–477 (1995).
24. Wu, J., Motto, D. G., Koretzky, G. A. & Weiss, A. Vav and SLP-76 interact and functionally cooperate in IL-2 gene activation. *Immunity* **4**, 593–602 (1996).
25. Segal, B. M., Dwyer, B. K. & Shevach, E. M. An interleukin (IL)-10/IL-12 immunoregulatory circuit controls susceptibility to autoimmune disease. *J. Exp. Med.* **187**, 537–546 (1998).
26. Hardt, W. -D., Chen, L. M., Schuebel, K. E., Bustelo, X. R. & Galan, J. E. S. typhimurium encodes an activator of Rho GTPase that induces membrane ruffling and nuclear responses in host cells. *Cell* **93**, 815–826 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank L. X. Zheng, Y. C. Liu and Y. Pewzner-Jung for the reagents, and K. Druey, D. Garboczi, R. N. Germain, W. E. Paul and R. H. Schwartz for critical comments on the manuscript.

Correspondence and requests for materials should be addressed to H.G. (e-mail: hgu@niaid.nih.gov).

Measurement of thermal contribution to photoreceptor sensitivity

Ari Koskelainen*, Petri Ala-Laurila*, Nanna Fyhrquist† & Kristian Donner†

* Laboratory of Biomedical Engineering, PO Box 2200, Helsinki University of Technology, FIN-02015 HUT, Finland

† Department of Biosciences, Division of Animal Physiology, FIN-00014 University of Helsinki, Finland

Activation of a visual pigment molecule to initiate phototransduction requires a minimum energy, E_a , that need not be wholly derived from a photon, but may be supplemented by heat¹. Theory^{2,3} predicts that absorbance at very long wavelengths declines with the fraction of molecules that have a sufficient complement of thermal energy, and that E_a is inversely related to the wavelength of maximum absorbance ($\lambda_{\max}$) of the pigment. Consistent with the first of these predictions, warming increases relative visual sensitivity to long wavelengths^{4–8}. Here we measure this effect in amphibian photoreceptors with different pigments to estimate E_a (refs 2, 5–7) and test experimentally the predictions of an inverse relation between E_a and $\lambda_{\max}$. For rods and 'red' cones in the adult frog retina, we find no significant difference in E_a between the two pigments involved, although their $\lambda_{\max}$ values are very different. We also determined E_a for the rhodopsin in toad retinal rods—spectrally similar to frog rhodopsin but differing in amino-acid sequence—and found that it was significantly higher.

In addition, we estimated E_a for two pigments whose $\lambda_{\max}$ difference was due only to a chromophore difference (A1 and A2 pigment, in adult and larval frog cones). Here E_a for A2 was lower than for A1. Our results refute the idea of a necessary relation between $\lambda_{\max}$ and E_a , but show that the A1 → A2 chromophore substitution decreases E_a .

Phototransduction is initiated when a molecule of visual pigment absorbs a photon of sufficient energy to isomerize the chromophore. Stiles² suggested that the energy barrier for pigment activation (E_a) does not set a sharp minimum requirement on photon energy (hc/λ), because part of the total energy can be drawn from appropriate vibrational modes of the molecule⁶. If this were not the case, every visual pigment would have a limiting wavelength $\lambda_a = hc/E_a$ where absorbance (or visual sensitivity) drops precipitously to zero, whereas all spectra in fact show a continuous, smooth decline. When plotted on logarithmic ordinates against $1/\lambda$ (thus photon energy), the decline follows a straight line at the longest wavelengths for as far as absorbance or sensitivity can be measured. This is predicted on the basis of classical statistical physics, if the probability that low-energy photons produce isomerizations is proportional to the cumulative distribution of pigment molecules on thermal energy levels greater than $E_a - hc/\lambda$. One corollary is that absorbances/sensitivities at long wavelengths ($\lambda > \lambda_a$) should be relatively increased by warming, which has indeed been observed in psychophysical^{4,6}, behavioural⁵ and electrophysiological^{7,8} measurements. The theory permits calculation of E_a from the size of the temperature effect at any wavelength $\lambda > \lambda_a$; the effect is translated into equivalent photon energy via the local slope of the sensitivity spectrum on a $1/\lambda$ abscissa (see equation (1) in Methods)^{4–7}.

We use this method to estimate E_a for spectrally different and spectrally similar visual pigments. The purpose is to test the idea of a straightforward relation between spectral properties and E_a , such that—for example—red-sensitivity would imply a lower energy barrier for activation than blue-sensitivity. This idea has influenced thinking on the ecology of visual pigments for almost half a century^{3,9–13} but has been difficult to test experimentally. In nature, the spectral properties of the visual pigment molecules are 'tuned' by at least two different molecular mechanisms: amino-acid substitutions in the protein (opsin) part¹⁴, acting on an evolutionary timescale, or a change of the prosthetic group (the chromophore) in one and the same opsin, effective on a physiological timescale¹⁵. In vertebrates, the latter mechanism entails a switch between retinaldehyde (retinal A1) and 3-dehydroretinaldehyde (retinal A2), as found in connection with (for example) seasonal or migratory changes in fishes, and in amphibian metamorphosis^{15,16}. The A1 → A2 substitution red-shifts the pigment in a regular manner¹⁷. As all known vertebrate visual-pigment spectra can be described over the main absorption range by either of two universal templates (A1 or A2) with $\lambda_{\max}$ as sole variable¹⁸, a supposed relation between spectral properties and the energy barrier for activation may to a first approximation be treated as a relation between $\lambda_{\max}$ and E_a .

Our primary goal required comparison between two pigments differing significantly in $\lambda_{\max}$ owing to different opsins. For this we studied the pigments of the main rod and cone populations in the

Table 1 Apparent activation energies for visual pigments in six amphibian photoreceptors

Source	Chromophore	$\lambda_{\max}$ (nm)	E_a (kcal mol ⁻¹)†	Number of points	Number of retinas
<i>Rana temporaria</i> rod	A1	502	45.7 ± 0.4	18	3
<i>Rana temporaria</i> cone	A1	562	45.5 ± 0.4	20	4
<i>Bufo bufo</i> rod	A1	502	49.2 ± 0.6**	21	3
<i>Rana temporaria</i> juv. cone	A2	629	(≤) 40.4 ± 1.6*	9	3
<i>Xenopus laevis</i> rod	A2	528	41.2 ± 1.5*	20	6

Values of the apparent activation energy, E_a , were estimated from equation (1). Statistical significances indicated for rods refer to a comparison with adult frog rods and for the cone A2 pigment to a comparison with its A1 pair. The amino acid sequences of the rod opsins of *Rana temporaria*, *Bufo bufo* and *Xenopus laevis* are compared in ref. 20.

† Values are mean ± s.e.m. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$.

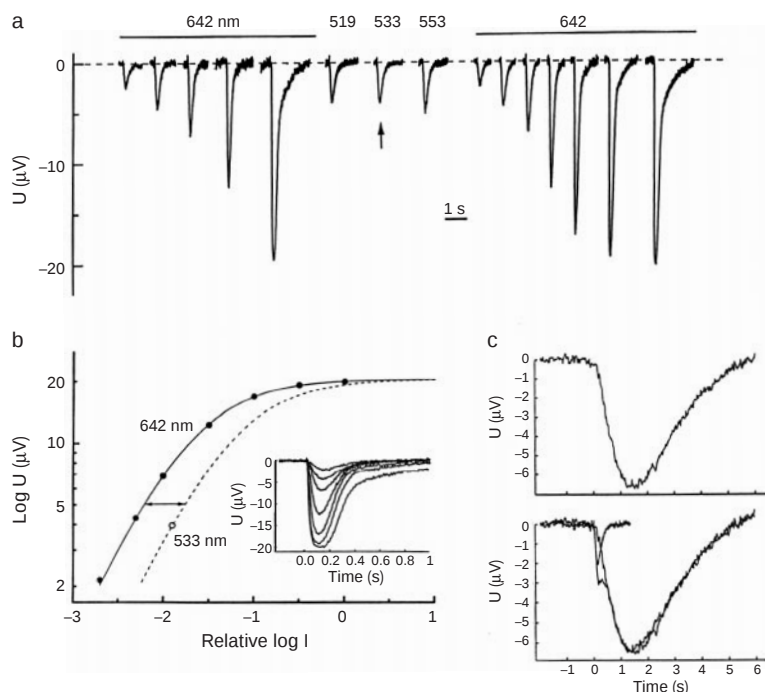


Figure 1 Measurement of spectral sensitivities of rods and cones in the isolated retina. **a**, The protocol for determining spectral sensitivities, exemplified by recordings from cones of the adult frog. Sequences of stimulation with five or seven flash intensities at a reference wavelength (642 nm for cones) were interleaved with series of 3–20 dim flashes at each of 3–5 different test wavelengths. Averaged dim-flash responses are shown for 519, 533 and 553 nm. The timescale bar (1 s) refers to the kinetics of the (arbitrarily spaced) responses. **b**, Response families (inset) recorded at the reference wavelength were used for construction of log (flash intensity) – log (response amplitude) functions (filled circles), described by a Michaelis curve (solid line). Sensitivity at each test

wavelength relative to the reference was determined from the required lateral shift of the curve (dashed line); here we show data for the response amplitude measured at 533 nm (open circle). **c**, Isolation of the rod component of mixed responses from the dark-adapted retina. The upper trace is a pure rod response, as obtained at shorter wavelengths (518 nm, single record). The lower trace is a mixed rod/cone response as obtained at long wavelengths (here 720 nm). The rod and cone components can be reliably separated by their kinetics, as shown by the superimposed traces: one reproduces the upper 518-nm trace and the other is a pure cone response to the same 720-nm flash intensity as for the mixed response, but recorded 4 s after a rod-saturating blue pre-flash¹⁹. Temperature, 25 °C.

adult frog retina, the ‘rhodopsin’ rods and the ‘red’ cones (*Rana temporaria*, A1, $\lambda_{\max} = 502$ and 562 nm, respectively)^{16–19}. Second, we wanted to compare two spectrally similar pigments with different opsins, and for this purpose we included the ‘frog-like’ rhodopsin of the toad, *Bufo bufo* (A1, $\lambda_{\max} = 502$ nm)^{18,20}. Third, to compare two pigments differing in $\lambda_{\max}$ due to a chromophore difference alone, we studied the A2 variant of the frog’s red-cone pigment ($\lambda_{\max} = 629$ nm), which occurs naturally in tadpoles¹⁶. Spectra were determined from electrophysiologically recorded spectral sensitivities of the respective photoreceptor cells (see Methods and Fig. 1), which is the only way to measure the far red end with high accuracy¹⁸, and has the further advantage that results refer to the functional visual pigment in reasonably natural conditions.

Figure 2a shows average spectral sensitivities of rods and red cones recorded in retinas of adult frogs at 5.4 °C (‘cold’, denoted C) and 25.0 °C (‘warm’, denoted W). In both photoreceptor types, relative W sensitivities rise successively above C sensitivities at the longest wavelengths. The sensitivity difference at each wavelength beyond the divergence in the vicinity of 640 nm (on the wavenumber scale, around $(1.6\text{--}1.5) \times 10^6 \text{ m}^{-1}$) yields one estimate of E_a (equation (1)). The mean $\pm$ s.e.m. of the estimates is $E_{ar} = 45.7 \pm 0.4 \text{ kcal mol}^{-1}$ for rods and $E_{ac} = 45.5 \pm 0.4 \text{ kcal mol}^{-1}$ for cones. These s.e.m.s include not only the variability between the point estimates, but also the uncertainty associated with the relative positioning of C and W spectra on the ordinate. The 95% confidence interval for the difference ($E_{ar} - E_{ac}$) is $0.2 \pm 1.8 \text{ kcal mol}^{-1}$; that is, only with probability $P < 0.05$ may E_{ac} be more than 4.4% smaller than E_{ar} . The hypothesis that the red shift from the rod to the red cone pigment is wholly due to lowered activation energy requires $E_a \propto 1/\lambda_{\max}$, implying that E_{ac} be 10.7% smaller than E_{ar} . This is rejected on the $P < 0.0005$ level.

Results from the spectrally similar rods of toad and frog are compared in Fig. 2b. When the two C spectra, as well as the shorter-wavelength limbs of the two W spectra, are optimally matched, the W data for toad at long wavelengths are seen to lie systematically above those for frog. This suggests that toad rhodopsin has a higher activation energy, and indeed the mean estimates differ significantly ($P < 0.01$; see Table 1). To summarize, two pigments with the same $\lambda_{\max}$ may have different E_a , and two pigments with similar E_a may have substantially different $\lambda_{\max}$. Equal activation energy appears to be neither a necessary nor a sufficient condition for equal $\lambda_{\max}$.

We show spectra of red cones of *Rana* tadpoles with about 40% A2 in Fig. 3. It may be seen that the temperature effect is much weaker than in Fig. 2. With the given proportion of A2, more than 90% of the threshold response at wavelengths $\lambda > 700$ nm is due to excitations in A2 molecules. Estimation of E_a based on these wavelengths gave $40.4 \pm 1.6 \text{ kcal mol}^{-1}$ for frog cone opsin with chromophore A2, clearly different from the value for its A1 counterpart ($P < 0.05$). However, the A2 estimate should properly be regarded as an upper bound (see Methods), and the true difference could in fact be greater. For comparison, we studied a spectrally similar A2 pigment in the red cones of adult clawed toad (*Xenopus laevis*, data not shown), finding hints of a temperature effect only from 752 nm (suggesting that $E_a \approx 39 \text{ kcal mol}^{-1}$). A third A2 pigment studied, the 528-nm porphyropsin of *Xenopus* rods (the pair of an A1 rhodopsin at ~ 502 nm), also gave a low E_a estimate ($41.2 \pm 1.6 \text{ kcal mol}^{-1}$), supporting the notion of a generic difference between A2- and A1-based pigments.

The results are summarized in Table 1. The activation energies for A1 pigments are in broad agreement with estimates from photo-calorimetry on bovine rhodopsin ($45\text{--}48 \text{ kcal mol}^{-1}$; ref. 21) as well as from bleaching studies on frog and bovine rhodopsin extracts,

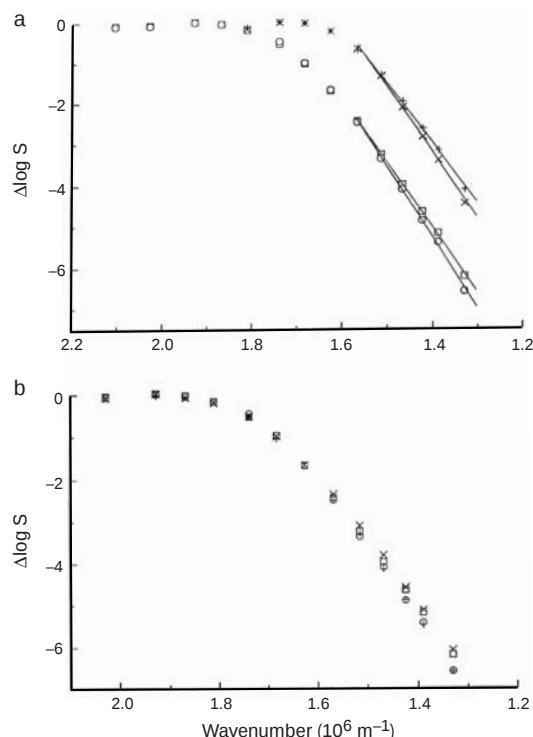


Figure 2 Effect of temperature on sensitivity spectra. **a**, Spectral sensitivities of rods and 'red' cones in the adult frog retina at 5.4 °C ('cold', C) and 25 °C ('warm', W). Symbols: circles, rod-C; squares, rod-W; crosses, cone-C; plus signs, cone-W. The rod spectra show averages from 3 retinas, the cone spectra averages from 4 retinas. Straight lines have been drawn in the long-wavelength region for visual guidance; regression coefficients calculated over the interval 638–752 nm are ($\times 10^{-6} \text{ m}$) 17.1 (rod-C), 15.8 (rod-W), 16.1 (cone-C) and 14.5 (cone-W). **b**, Comparison of spectral sensitivities of frog and toad rods. Frog data (circles and squares) are reproduced from **a**. Toad data (averages from 3 retinas) are marked by plus signs (C) and crosses (W). Linear regression coefficients calculated for the toad data over the interval 638–752 nm are $17.1 \times 10^{-6} \text{ m}$ (6.9 °C) and $15.9 \times 10^{-6} \text{ m}$ (25.8 °C) (lines not shown).

whether purely thermal (44 kcal mol⁻¹; ref. 22) or involving combinations of light and heat (44–48.5 kcal mol⁻¹; ref. 1). Lower E_a in A2 pigments is expected *a priori* from the extra C=C bond in the π -electron system of 3-dehydroretinal, and holds as a qualitative rule for pigments in solution²³.

The thermal-energy contribution to visual excitation is functionally important in two respects. First, as shown above, it expands the range of long-wavelength light available for vision. Our results suggest that thermal extension of the spectrum at the red end can be achieved in two ways: either by decreasing the activation energy (as by A1 → A2 substitution), or by increasing the capacity to recruit thermal energy (vibrational modes) towards chromophore isomerization⁶. Second, it implies that pigments may be activated even in the absence of photons. This will cause an intrinsic background 'light' of randomly occurring quantal excitations ('dark events'), which (as they are identical to those due to photons) constitute a type of noise that must necessarily degrade the detectability of real light^{24,25}. Improving red-sensitivity by increasing the thermal-energy contribution (regardless by which means) is likely to increase the susceptibility to purely thermal activation. Noise reduction has been proposed to explain the widespread occurrence of visual pigments (or spectral sensitivities) that are blue-shifted relative to the ambient illumination^{9–13}, including the Purkinje shift from photopic (daylight) vision to relatively short-wavelength-sensitive scotopic (night) vision³. In rods, a general correlation between red-sensitivity and high rates of dark events has been

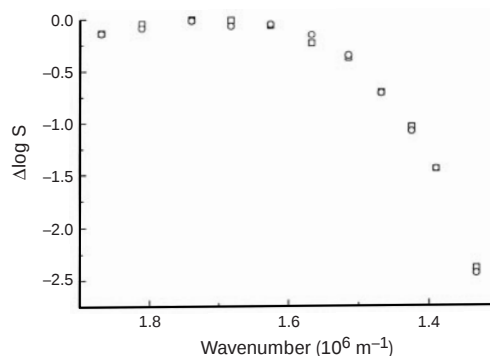


Figure 3 Spectral sensitivities of frog tadpole 'red' cones (circles, 7.9 °C; squares, 24 °C). Averages from 3 retinas. The local slope between the last two points (720–752 nm) is $16.8 \times 10^{-6} \text{ m}$.

experimentally demonstrated^{11,12}; this evidence mainly rests on spectral variation associated with chromophore differences. In cones, which offer a wider range of opsin-based spectral variation, the 'dark' noise power of single red cones^{26,27}, as well as the variability of light detection by the dark-adapted human L + M cone system²⁸, are consistent with high rates of dark events. (Quantal excitations in cones are too small to be directly recorded.)

We note, however, that our estimates of E_a for A1-based pigments (Table 1) are more than twice those obtained from the temperature dependence of dark-event rates in toad rhodopsin rods²⁵. This is consistent with the idea that purely thermal activation occurs by a different molecular route than does activation by photons, even when the latter depends on support from thermal energy²⁹. □

Methods

Preparation and recording

Mass photoresponses of rods and cones were recorded as electroretinogram potentials across the isolated retina, superfused and illuminated from the receptor side. Synaptic transmission was blocked by aspartate. Spectral sensitivities were measured with interference filters (Schott DIL or Melles Griot) selected from a complement of 29 individually calibrated filters with peak transmissions covering the range 397–802 nm (in most cases, however, the lamp output and spectral sensitivities of the cells restricted the useful wavelength range to $\leq 752 \text{ nm}$). In each experiment at least two spectra were recorded, one ('C') usually at a temperature in the interval 5–8 °C and one ('W') usually at 25–27 °C; the order of W and C was varied between experiments. Temperature was monitored with a thermistor in the bath close to the retina. After dissection and after each temperature change, the retina was allowed to adapt for at least 1 h before recording began. The protocol is shown in Fig. 1. At each wavelength, responses to 3–20 nominally identical 20-ms flashes of dim light were averaged. Possible changes in saturating response amplitude during the experiment were corrected for by linear interpolation between interleaved intensity-response functions recorded with a reference wavelength. Cone responses were isolated either by selective suppression of rods with steady background light (in two of four frog cone experiments in Table 1), or, in the dark-adapted retina, by stimulation at a fixed time after a strong blue pre-flash, while rods were still saturated but cones had essentially recovered¹⁹ (two experiments). Rod responses were recorded in the dark-adapted retina; cone contributions to responses to longer wavelengths were eliminated on the basis of kinetics (Fig. 1c). Distortion by selective absorption in structures other than visual pigment, waveguide properties of photoreceptors, and response contributions from shorter-wavelength-sensitive cone and rod types can be ruled out in the wavelength range used here (near and beyond the peak of rhodopsin rods and 'red' cones), and with the present recording geometry¹⁹. However, rhodopsin absorbance will affect the shape of both rod and cone spectra, as a virtually full dark-adapted complement of rhodopsin was present throughout all experiments. This will both broaden the peak of rod spectra compared with dilute pigment ('self-screening' with optical density ~ 0.5)^{16,19}, and depress apparent cone sensitivities on the short-wavelength side of the peak¹⁹. But screening by rhodopsin did not change during experiments, as there was no substantial pigment loss by bleaching. In the 'worst' case (cone experiments with a continuous rod-depressing background), the bleaching velocity was 1% per 30 min; the isolated retina is known to possess a sufficient capacity for pigment regeneration to compensate for this completely over at least 5 h. Screening is unimportant in the long-wavelength range, where rhodopsin absorbance is low. See refs 19 and 30 for further technical details.

Analysis

At each wavelength, the log sensitivity (log S) difference compared with a reference wavelength was determined. All differences were expressed relative to log peak sensitivity ($= 0$) and these normalized values of log S for each wavelength were averaged across

experiments for each photoreceptor type, C and W separately. The optimal relative positioning along the logS axis of the averaged C and W spectra for each photoreceptor type was found by minimizing the sum of squares (SS) of the differences between C and W points at 'shorter' wavelengths (below hc/E_a , determined by iteration). The least SS, denoted SS_{μ} , provides directly a statistical measure of the 'matching' uncertainty connected with the vertical alignment (see below). The sensitivity values of the averaged spectra under the optimal match are denoted $\Delta \log S$. The difference $\Delta \log S_{W_i} - \Delta \log S_{C_i}$ recorded at each 'long' wavelength λ_i was converted into its photon energy equivalent through the equation^{2,5-7}.

$$hc/\lambda_a = E_a = hc/\lambda_i + (hc/T)[- \partial \log S / \partial (1/T)]_i / [\partial \log S / \partial (1/\lambda)]_i \quad (1)$$

For the calculations, $[- \partial \log S / \partial (1/T)]_i = (\Delta \log S_{W_i} - \Delta \log S_{C_i}) / (1/T_C - 1/T_W)$, $T = T_C$, and $[\partial \log S / \partial (1/\lambda)]_i$ is the local slope of the C-spectrum at λ_i , taken as the mean of $(\Delta \log S_{C(i-1)} - \Delta \log S_{C(i)}) / (1/\lambda_{i-1} - 1/\lambda_i)$ and $(\Delta \log S_{C_i} - \Delta \log S_{C(i+1)}) / (1/\lambda_i - 1/\lambda_{i+1})$. (For the largest λ_i , the slope from λ_{i-1} to λ_i was accepted as such.)

The statistical error of the estimate depends on two variance components. (1) Variance connected with the overall vertical matching of C and W spectra was estimated by a 'matching' mean square $s_{\mu}^2 = A_{\max} SS_{\mu} / (m - 1)$. Here $A_{\max}$ is the largest value of the energy conversion factor $hc / [(1 - T_C/T_W) \partial \log S_C / \partial (1/\lambda)]_i$ in that data set (corresponding to the smallest local slope), SS_{μ} is as defined earlier, and m is the number of (shorter-wavelength) data point pairs used for the matching. (2) Variance between the point estimates of E_a under a fixed C-W match was estimated by a conventional 'sample' mean square $s_p^2 = SS_p / (M - 1)$. Here, SS_p is the sum of squares of the sample and M is sample size, that is, the number of (longer-wavelength) data point pairs used for estimation. Since C-W matching and E_a estimation were based on non-overlapping data sets, the estimator of total variance (denoted s_{Σ}^2) is a sum of s_{μ}^2 and s_p^2 , weighted by their respective degrees of freedom ($df_{\mu} = m - 1$ and $df_p = M - 1$). The total s.e.m. is s_{Σ} / M . Confidence limits and statistical significance testing were based on Student's t -test with degrees of freedom according to the appropriate variance measure in each case.

The validity of equation (1) depends on the existence of the true positive differences $\Delta \log S_W - \Delta \log S_C$. Otherwise (for random, zero or negative differences) no definite values for E_a can be obtained, only an upper limit ($E_a \leq hc/\lambda_i$). In the case of frog tadpole cones only three wavelengths were available for estimation (Fig. 3), and there is the possibility that the small differences observed reflect essentially random variation; hence, the corresponding estimate in Table 1 should be regarded as an upper bound. In *Xenopus* red cones, temperature effects were even less reliable.

For determining the A1/A2 ratio in A2-containing photoreceptors, a curve-fitting program was used where given proportions of A1 and A2 templates (recently tuned for best fit to a large body of spectral data)¹⁸ are added together, while the $\lambda_{\max}$ difference within an A1-A2 pair is constrained to obey the Dartnall-Lythgoe relation¹⁷. (Description of the shape of our spectra with templates for absorbance of dilute pigment requires rhodopsin screening be taken into account; see above.) In our adult *Xenopus*, we detected no A1 chromophore, while the *Rana* tadpoles were found to have approximately a 0.6/0.4 mixture of A1 and A2. The program was also used to analyse the expected contribution of the 562-nm A1 component to the sensitivity of these tadpole cones at different wavelengths. Only the longest wavelengths, with less than 8% contribution from A1, were used in determining E_a for the A2 pigment.

Received 16 March; accepted 2 November 1999.

1. St George, R. C. C. The interplay of light and heat in bleaching rhodopsin. *J. Gen. Physiol.* **35**, 495-517 (1952).
2. Stiles, W. S. in *Trans. of the Optical Convention of the Worshipful Company of Spectacle Makers* 97-107 (Spectacle Makers' Co., London, 1948).
3. Barlow, H. B. Purkinje shift and retinal noise. *Nature* **179**, 255-256 (1957).
4. de Vries, H. Der Einfluss der Temperatur des Auges auf die spektrale Empfindlichkeitskurve. *Experientia* **4**, 357-358 (1948).
5. Denton, E. J. & Pirenne, M. H. The visual sensitivity of the toad *Xenopus laevis*. *J. Physiol.* **125**, 181-207 (1954).
6. Lewis, P. R. A theoretical interpretation of spectral sensitivity curves at long wavelengths. *J. Physiol.* **130**, 45-52 (1955).
7. Srebro, R. A thermal component of excitation in the lateral eye of *Limulus*. *J. Physiol.* **187**, 417-425 (1966).
8. Lamb, T. D. Effects of temperature changes on toad rod photocurrents. *J. Physiol.* **346**, 557-578 (1984).
9. Lythgoe, J. N. in *Sensory Biology of Aquatic Animals* (eds Atema, J., Fay, R. R., Popper, A. N. & Tavolga, W. N.) 57-82 (Springer, New York, 1988).
10. Goldsmith, T. H. in *Facets of Vision* (eds Stavenga, D. G. & Hardie, R. C.) 1-14 (Springer, Berlin, 1989).
11. Donner, K., Firsov, M. L. & Govardovskii, V. I. The frequency of isomerization-like "dark" events in rhodopsin and porphyropsin rods of the bullfrog retina. *J. Physiol.* **428**, 673-692 (1990).
12. Firsov, M. L. & Govardovskii, V. I. Dark noise of visual pigments with different absorption maxima. *Sensornyye Sistemy* **4**, 25-34 (1990). (In Russian).
13. Bowmaker, J. K. et al. Visual pigments and the photic environment: The cottoid fish of Lake Baikal. *Vision Res.* **34**, 591-606 (1994).
14. Neitz, M., Neitz, J. & Jacobs, G. H. Spectral tuning of pigments underlying red-green color vision. *Science* **252**, 971-974 (1991).
15. Bridges, C. D. B. in *Handbook of Sensory Physiology* Vol. VII/1 (ed. Dartnall, H. J. A.) 417-480 (Springer, Berlin, 1972).
16. Reuter, T. Visual pigments and ganglion cell activity in the retinae of tadpoles and adult frogs (*Rana temporaria* L.). *Acta Zool. Fenn.* **122**, 1-64 (1969).
17. Dartnall, H. J. A. & Lythgoe, J. N. The spectral clustering of visual pigments. *Vision Res.* **5**, 81-100 (1965).
18. Govardovskii, V. I., Fyhrquist, N., Reuter, T., Kuzmin, D. G. & Donner, K. In search of the visual pigment nomogram. *Vis. Neurosci.* (in the press).

19. Koskelainen, A., Hemilä, S. & Donner, K. Spectral sensitivities of short- and long-wavelength sensitive cone mechanisms in the frog retina. *Acta Physiol. Scand.* **152**, 115-124 (1994).
20. Fyhrquist, N. et al. Rhodopsins from three frog and toad species: sequences and functional comparisons. *Exp. Eye Res.* **66**, 295-305 (1998).
21. Cooper, A. Energy uptake in the first step of visual excitation. *Nature* **282**, 531-533 (1979).
22. Lythgoe, R. J. & Quilliam, J. P. The thermal decomposition of visual purple. *J. Physiol.* **93**, 24-38 (1938).
23. Williams, T. P. & Milby, S. E. The thermal decomposition of some visual pigments. *Vision Res.* **8**, 359-367 (1968).
24. Barlow, H. B. Retinal noise and absolute threshold. *J. Opt. Soc. Am.* **46**, 634-639 (1956).
25. Baylor, D. A., Matthews, G. & Yau, K.-W. Two components of electrical dark noise in toad retinal rod outer segments. *J. Physiol.* **309**, 591-621 (1980).
26. Lamb, T. D. & Simon, E. J. Analysis of electrical noise in turtle cones. *J. Physiol.* **272**, 435-468 (1977).
27. Schnapf, J. L., Nunn, B. J., Meister, M. & Baylor, D. A. Visual transduction in cones of the monkey *Macaca fascicularis*. *J. Physiol.* **427**, 681-713 (1990).
28. Donner, K. Noise and the absolute thresholds of cone and rod vision. *Vision Res.* **32**, 853-866 (1992).
29. Barlow, R. B., Birge, R. R., Kaplan, E. & Tallent, J. R. On the molecular origin of photoreceptor noise. *Nature* **366**, 64-66 (1993).
30. Donner, K., Hemilä, S. & Koskelainen, A. Light adaptation of cone photoresponses studied at the photoreceptor and ganglion cell levels in the frog retina. *Vision Res.* **38**, 19-36 (1998).

Acknowledgements

We thank V. I. Govardovskii for the nomogram-fitting program used for spectral analysis of pigments, H. Rita for statistical advice, C. Haldin and S. Pietilä for technical assistance, T. Reuter for comments on the manuscript, and Kilpisjärvi Biological Station for providing experimental animals. This work was supported by the Academy of Finland and by the Finnish Graduate Schools of Neuroscience and of Molecular Nanotechnology.

Correspondence and requests for materials should be addressed to A.K.
(e-mail: ari.koskelainen@hut.fi).

Bioorganic synthesis of lipid-modified proteins for the study of signal transduction

Benjamin Bader*, Karsten Kuhn*†, David J. Owen*†‡, Herbert Waldmann*†, Alfred Wittinghofer* & Jürgen Kuhlmann*

* Max-Planck Institut für Molekulare Physiologie, Otto Hahn Strasse 11, 44227 Dortmund, Germany

† Universität Dortmund, FB 3, Organische Chemie, Otto Hahn Strasse 11, 44227 Dortmund, Germany

Biological membranes define the boundaries of the cellular compartments in higher eukaryotes and are active in many processes such as signal transduction and vesicular transport. Although post-translational lipid modification of numerous proteins in signal transduction is crucial for biological function¹, analysis of protein-protein interactions has mainly focused on recombinant proteins in solution under defined *in vitro* conditions. Here we present a new strategy for the synthesis of such lipid-modified proteins. It involves the bacterial expression of a carboxy-terminally truncated non-lipidated protein, the chemical synthesis of differently lipidated peptides representing the C terminus of the proteins, and their covalent coupling. Our technique is demonstrated using Ras constructs, which exhibit properties very similar to fully processed Ras, but can be produced in high yields and are open for selective modifications. These constructs are operative in biophysical and cellular assay systems, showing specific recognition of effectors by Ras lipoproteins inserted into the membrane surface of biosensors and transforming activity of oncogenic variants after microinjection into cultured cells.

Ras, a GTP-binding protein involved in signal transduction², has to undergo several post-translational lipid-modification steps as a crucial prerequisite for biological activity³. The most important

‡ Current address: Department of Medicinal Chemistry, Victorian College of Pharmacy, 381 Royal Parade, Parkville, Victoria 3052, Australia.

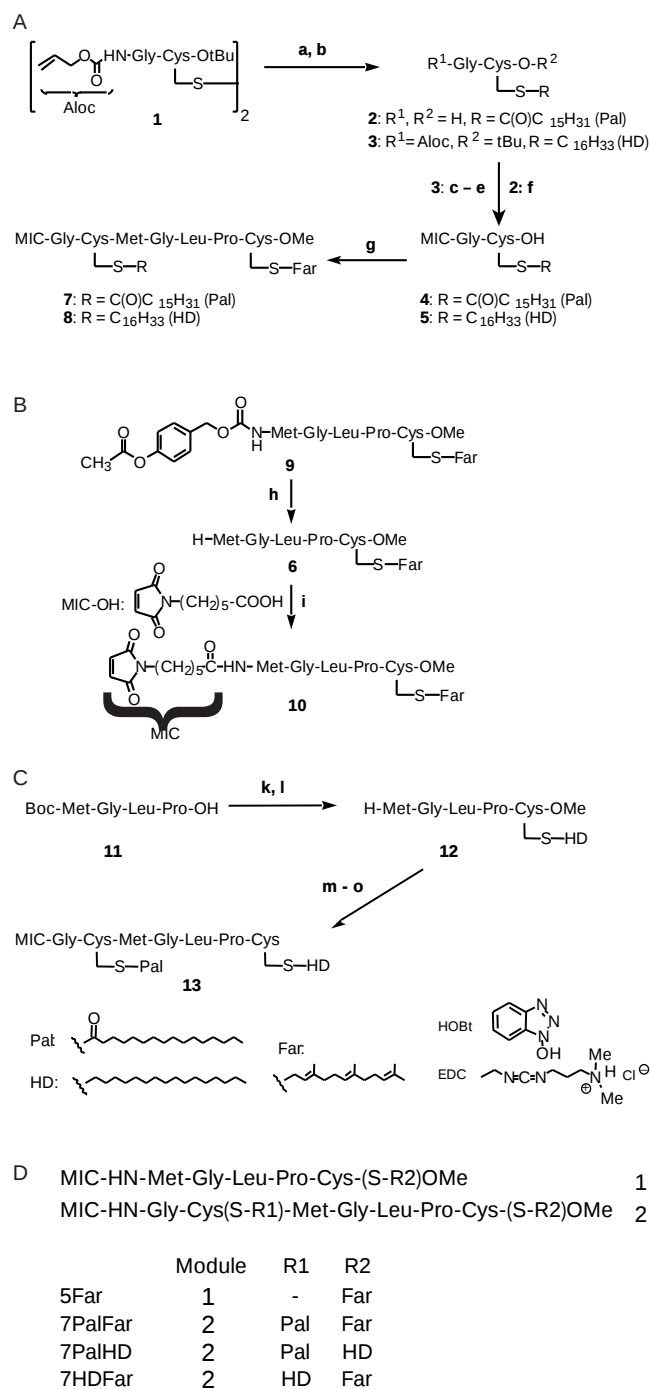


Figure 1 Synthesis of differently lipidated Ras peptides. See Methods for details. **A**, **a**, Dithiothreitol (DTT); **b**, $\text{C}_{16}\text{H}_{33}\text{Br}$, DMF, 0°C , 5d, **3**, 60% (two steps); **c**, 5 mol% $\text{Pd}(\text{PPh}_3)_4$, 5 equiv. dimethylmalonate, THF, room temperature; **d**, 1 equiv. maleimidocaproic acid (MIC-OH), 1 equiv. HOBt, 1 equiv. EDC, CH_2Cl_2 , 51% (two steps); **e**, CF_3COOH , thioanisole, quant.; **f**, 1 equiv. MIC-OH, 1 equiv. HOBt, 1 equiv. EDC, CH_2Cl_2 , 62%; **g**, **6**, 1 equiv. HOBt, 1 equiv. EDC, CH_2Cl_2 , **7**, 48%; **8**, 75%. for the synthesis of **2** via selective removal of the allyl ester protecting group see ref. 7. **B**, **h**, Lipase from *Mucor miehei*, 0.2M KI buffer, 20 vol.% CH_3OH , pH 5, 100 equiv. NaHS, 30°C , 48%; **i**, 1 equiv. MIC-OH, 1 equiv. HOBt, 1 equiv. EDC, CH_2Cl_2 , 78%. **C**, **k**, H-Cys(HD)-OMe, 1 equiv. HOBt, 1 equiv. EDC, CH_2Cl_2 , 73%; **l**, CF_3COOH , thioanisole, quant.; **m**, Boc-Gly-Cys(Pal)-OH, 1 equiv. HOBt, 1 equiv. EDC, CH_2Cl_2 , 61%; **n**, CF_3COOH , thioanisole; **o**, 1 equiv. MIC-OH, 1 equiv. HOBt, 1 equiv. EDC, CH_2Cl_2 , 63% (two steps). **D**, Nomenclature of lipopeptides.

events of the Ras pathway occur at the plasma membrane and inhibition of the membrane insertion of Ras is a primary target for the development of anti-Ras drugs⁴. However, most biophysical studies on protein–protein interactions involved in this scenario have been carried out with bacterially synthesized proteins lacking C-terminal modification. The isolation of completely modified proteins (for example, from baculoviral expression in SF9 cells) results in low yield with the problem of pronounced lability of thioester bonds throughout the purification steps⁵. Therefore, alternative techniques for the generation of modified proteins could provide new insight into protein–membrane interaction or the influence of lipid modification upon biological function. The approach presented here combines new strategies of organic synthesis in the building of complex lipopeptide structures with bacterial expression of proteins. We decided to use lipopeptides activated at the amino-terminal end with a maleimido modification as a sulphhydryl-group-specific electrophilic reagent. The linkage with the protein should be realized by conjugation with a truncated Ras protein carrying a C-terminal cysteine group.

The synthesis of lipidated Ras-peptides is severely complicated by their pronounced acid- and base-lability. Under acidic conditions the double bonds of the farnesyl group are attacked and at pH >7 the thioesters spontaneously hydrolyse. This sensitivity calls for the application of blocking groups that can be removed under the mildest conditions. By combining the use of enzymatic protecting group techniques with noble-metal catalysed removal of blocking functions, maleimido-modified Ras peptides embodying different lipid groups can efficiently be synthesized (Fig. 1). Maleimidocaproyl (MIC)-modified S-palmitoylated or S-hexadecyl-modified N-terminal dipeptides **4** and **5** were built up by employing the Pd(0)-catalysed removal of the allyloxycarbonyl (Aloc) urethane protecting group and the corresponding allyl ester via noble-metal mediated allyl transfer to a C-nucleophile as the key step⁶. After introduction of the maleimidocaproyl residue and selective C-terminal deprotection the resulting building blocks (**4** and **5**) were coupled with an S-farnesylated pentapeptide building block (**6**) to deliver maleimido-modified palmitoylated and farnesylated or hexadecylated and farnesylated heptapeptides MIC7PalFar (**7**) and MIC7HDFar (**8**) in high overall yields (Fig. 1a). The selectively N-deprotected pentapeptide building block (**6**) employed in this sequence was built up by enzyme-induced fragmentation of the AcOZ-urethane blocking function from masked farnesylated peptide (**9**) (ref. 7). Coupling of **6** with maleimidocaproic acid delivered farnesylated pentapeptide MIC5Far (**10**) (Fig. 1b). Palmitoylated and hexadecylated heptapeptide MIC7PalHD (**13**) was obtained from a tetrapeptide (**11**) by means of two successive chain elongation/N-terminal deprotection steps and final introduction of the MIC group (Fig. 1c).

The protecting group techniques employed in the syntheses detailed above proceed under exceptionally mild conditions and are very efficient. The acid and base labile side-chain modifications are not attacked at all during the removal of the blocking functions and the desired maleimido-modified lipidated peptides are efficiently obtained on a large scale if required, in gram amounts.

Truncated H-Ras protein with a C-terminal cysteine at position 181 was expressed recombinantly in *Escherichia coli*. Removing the C-terminal MSCKCVLS sequence had no effect on protein expression in *E. coli* CK600K (ref. 8), nucleotide binding and interaction with exchange factor Cdc25 or the Ras-binding domain (RBD) of Raf-kinase (data not shown). These Ras proteins were allowed to react with MIC-modified lipopeptides in stoichiometric amounts. Reaction products were purified by extraction⁹ and ion-exchange chromatography.

In addition to Cys 181 H-Ras has three (natural) cysteines, residues 51, 80 and 118, which could potentially also be modified, although only Cys 118 appears to be exposed to the solvent in the structure. Analysis of the reaction products by mass spectroscopy, both without and after protease digestion, confirmed that the C-terminal

cysteine was almost exclusively modified (data not shown). Regardless of the introduction of hydrophobic residues all Ras proteins retained high solubility in aqueous buffer and could be stored at -70°C without loss of activity. Only the long-time stability was somewhat reduced compared with non-modified Ras.

Artificial membranes can be generated by adsorption of lipid vesicles upon the surface of hydrophobic sensor chips as used in surface plasmon resonance (SPR) systems¹⁰. The polar side of the lipid phase is then exposed towards the reaction chamber, thereby mimicking the inner side of the plasma membrane. Hybrid lipoproteins spontaneously insert into the artificial bilayer when applied as aqueous solutions. The yield and stability of membrane insertion correlated with the degree of hydrophobic modification (Fig. 2a). This is in agreement with findings for biotinylated lipopeptides¹¹.

Ras interacts with several downstream effectors such as the protein kinases of the Raf family. Ras proteins form high-affinity complexes with the RBD of Raf-kinase only in their GTP-bound state¹². Incubation of RBD with GDP-bound Ras constructs inserted into the sensor surface showed only minor non-specific adsorption of RBD, whereas Ras in the GTP-bound state resulted in a specific increase in RBD binding. In contrast, an RBD mutant (R89L) that is deficient in Ras binding produced the same curvature for both nucleotides (Fig. 2b, net traces). Hence, hybrid Ras lipoproteins inserted in artificial lipid bilayers behave similarly to the non-modified protein immobilized in the hydrogel of a standard sensor chip¹³.

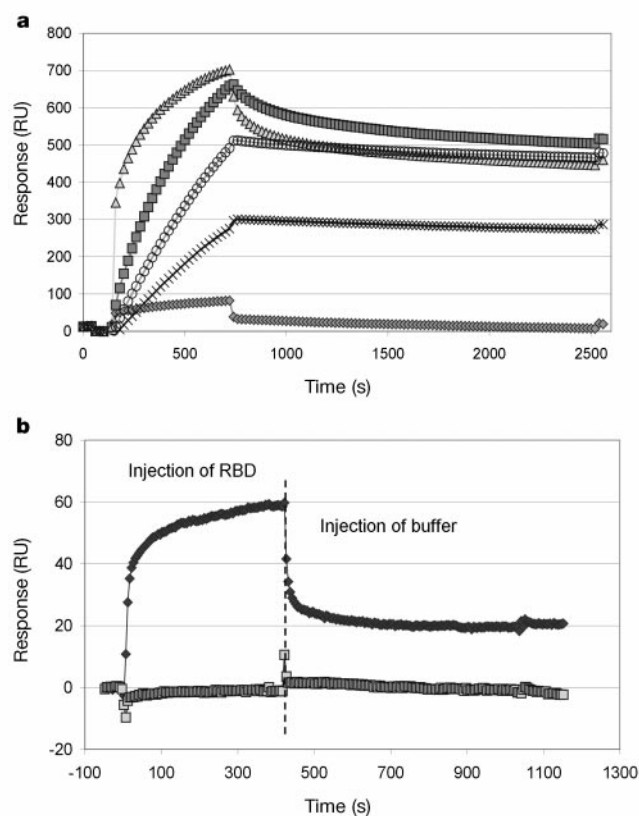


Figure 2 Insertion of modified H-Ras into a POPC-monolayer and binding of the cRaf-RBD. See Methods for details. **a**, Binding and dissociation of H-Ras proteins to the POPC-monolayer. Soluble H-RasΔC (diamonds), H-Ras coupled to MIC-5Far (triangles), MIC-7PalFar (circles), MIC-7HDFar (crosses) and *in vitro* farnesylated, full length H-Ras (squares) were applied to the preformed phospholipid layer. After 10 min protein solutions were replaced by SPR-buffer and signal decrease was observed for 30 min. **b**, Binding of cRaf-RBD to H-Ras-7HDFar inserted into the lipid monolayer. Net kinetics (binding of RBD to Ras*GppNHp-binding to Ras*GDP) were calculated for 250 nM of wild-type RBD (diamonds) and the mutant RBD^{R89L} (squares), which is deficient in Ras binding. Start and end of RBD injection are marked.

Activating mutations in Ras genes can contribute to either cell transformation and cancer development or differentiation of cells. The latter is the case for the rat pheochromocytoma cell line PC12 which differentiates very efficiently upon microinjection of oncogenic Ras proteins^{14,15}. After microinjection of full-length recombinant oncogenic Ras(G12V) protein, 79% of injected cells developed neurites (Fig. 3a and f). Full-length Ras protein is expected to be modified after microinjection in the same way as the intrinsic polypeptide chain. Indeed, if the truncated Ras(G12V)ΔC181 protein lacking both the C-terminal CaaX motif for isoprenylation and the first of two possible palmitoylation sites is microinjected, no neurites are formed (Fig. 3b and f). After injection of the coupling product of Ras(G12V)ΔC181 with the farnesylated and palmitoylated heptapeptide (MIC-7PalFar), a strong differentiating phenotype of the PC12 cells is induced (Fig. 3d) with comparable biological activity to the non-modified full-length protein (Fig. 3f). Hybrid oncogenic Ras with a farnesyl group but lacking a cysteine residue for palmitoylation is almost completely inactive in the assay (Fig. 3c and f). This is in agreement with transfection experiments using Ras constructs with mutations in the palmitoylatable cysteine

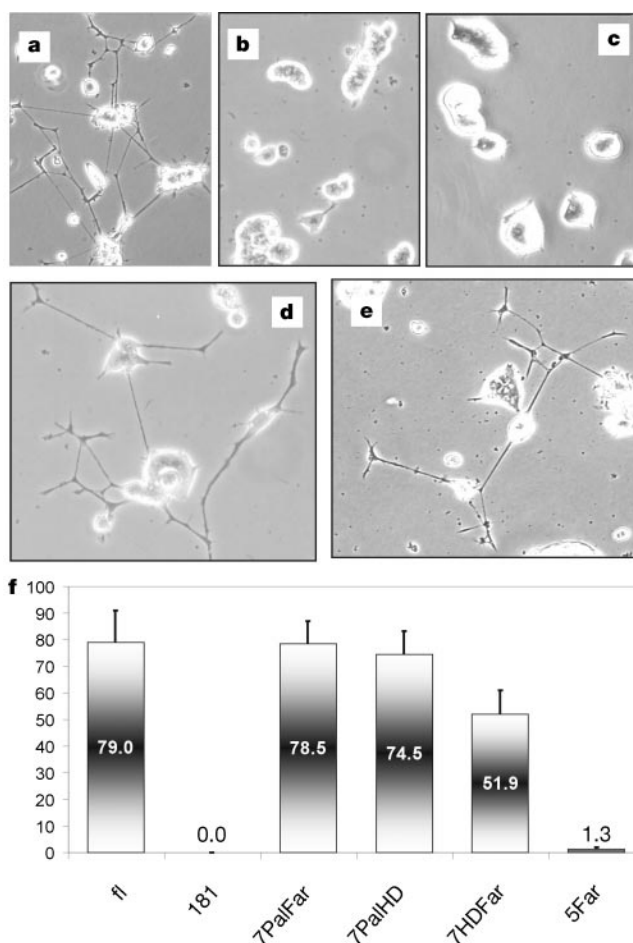


Figure 3 Transformation of PC12 cells by chemically coupled H-RasG12V. **a**, H-RasG12V full length (fl). **b**, C-terminally truncated H-RasG12VΔC 181. **c**, H-RasG12VΔC chemically coupled to farnesylated pentapeptide MIC-5Far. **d**, Farnesylated and palmitoylated heptapeptide (MIC-7PalFar). **e**, Heptapeptide MIC-7PalHD (farnesyl moiety replaced by a hexadecyl group). **f**, Quantification of the microinjection experiments. Transforming efficiencies of full length H-RasG12V (fl, 79.0% ± 14.0%), chemically coupled H-RasG12V-7PalFar (78.5% ± 8.0%), H-RasG12V-7PalHD (74.5% ± 9.0%), H-RasG12V-5Far (1.3 ± 2.5%) and H-RasG12V-7HDFar (51.9 ± 9.2%). Truncated H-RasG12VΔC without modification (181) induced no neurite-like outgrowths. Total numbers of injected cells in at least three independent experiments for each protein were 106, 66, 93, 47, 152 and 106 for H-RasG12V-fl, RasΔC, -7PalFar, 7PalHD, 5Far, and 7HDFar, respectively.

residues, which had no effect on farnesylation but markedly reduced transforming activity and plasma membrane localization¹⁶. Therefore one hydrophobic modification is not sufficient for the biological activity of H-Ras.

For one of the constructs we replaced the farnesyl thioether by an unbranched hexadecyl thioether. The corresponding coupling product shows the same pattern in neurite outgrowth as the farnesylated homologue (Fig. 3a and f). Therefore, no specific isoprenylation receptor has to be invoked to localize Ras to the plasma membrane¹⁷ and the biological activity of Ras is not dependent upon a farnesyl residue—a hydrophobic alkane chain and a palmitate are sufficient. If the labile palmitoyl thioester is replaced by a stable hexadecyl thioether the hybrid protein Ras(G12V)-7HDFar is also biologically active, but the efficiency of neurite formation is significantly reduced (Fig. 3f). This result may seem, at first, surprising. A non-hydrolysable thioether should strengthen the insertion of the Ras hybrid in the plasma membrane due to the reduced turnover compared to the natural thioester.

Nevertheless, the reduced biological readout of the 7HDFar-hybrid is an additional indication that the palmitoylation step of H-Ras occurs in the plasma membrane. The farnesylated and hexadecylated protein will insert into any cellular membrane with a low probability of detaching from it¹⁸. In contrast, the reversibility of thioester formation may allow palmitoylated constructs to switch between several membrane structures after induced or spontaneous hydrolysis of the thioester bond. If a corresponding prenyl protein-specific palmitoyltransferase was located exclusively in the plasma membrane, Ras constructs that can be palmitoylated would enrich there automatically.

We have shown that Ras proteins can be chemically modified with lipopeptides such that these hybrid proteins can insert into artificial and biological membranes. The neo-lipoproteins prove to be efficient tools for biochemical, biophysical and biological experiments. Large amounts of such hybrid proteins can be synthesized. The approach presented here uses an artificial non-peptide linker between the protein moiety and the peptide. It embodies a thioether consisting of five methylene groups and a maleimide ring instead of the natural peptide linkage. However, we have so far been unable to detect any significant influence of this artificial connection on *in vitro* interaction with effector proteins or on *in vivo* cellular response. In principle, the method presented here is applicable for many of the Ras-related GTP-binding proteins or the γ -subunit of heterotrimeric G proteins. □

Methods

Lipopeptide synthesis

Hexadecyl thioethers were synthesized by treatment of the corresponding disulphide with dithiothreitol (DTT) and alkylation of the liberated thiol with hexadecyl bromide. Removal of the Alloc group and the allyl ester function was achieved by treatment with Pd(PPh₃)₄ and dimethylmalonate as accepting C-nucleophile. Boc groups and tert-butyl esters were cleaved with CF₃COOH in the presence of thioanisole. Amide bonds were formed by activating the carboxylic acids with N-ethyl-N'-dimethylaminopropylcarbodiimide hydrochloride (EDC) and N-hydroxybenzotriazole (HOBt). The enzyme-labile 4-acetoxymethylbenzylcarboxyl (AcOZ) group was removed by treatment with lipase from *Mucor miehei* at pH 5 and in the presence of KI and NaSH as trapping reagents for the intermediary formed, quinone methide.

Protein synthesis

C-terminally truncated H-Ras for coupling was generated by introducing two stop codons at positions 182 and 183 of the pTac-Ras vector⁸ by polymerase chain reaction. Protein expression and purification of Ras and RBD proteins was performed as described^{8,12}.

Generation of hybrid proteins

The coupling reaction was performed with stoichiometric amounts of peptide and protein. The peptide was first dissolved in methanol and, after addition of 11% Triton X114 (polyethylene glycol tert-octylphenylether, Fluka), sonicated for 5 min. Finally, the same volume of protein solution in a buffer containing 20 mM Tris, 5 mM MgCl₂, pH 7.4 was added. The coupling mixture was covered with argon and incubated at 4°C for 16 h. For purification the hydrophobic coupling product was extracted by Triton X114 phase separation⁹. The protein extract was diluted to 2% Triton and applied on a DEAE-sepharose column. After washing with 20 mM Tris, 5 mM MgCl₂, 2 mM DTE pH 7.4 the

bound Ras protein was eluted with a sodium chloride gradient and concentrated. The product was analysed by mass spectrometry and SDS-PAGE.

SPR analysis

The hydrophobic surface of an HPA sensorchip (BIAcoreTM, Uppsala, Sweden) was loaded with 0.5 mM POPC (palmitoyl-oleoyl-phosphatidylcholine) vesicles, prepared by extrusion through a 50-nm polycarbonate filter^{19,20}. After a 1-min pulse of 10-mM NaOH to remove multiple layers of phospholipid, BSA (500 nM) was injected to cover bad spots on the sensor membrane. H-Ras constructs were applied with 50 ng ml⁻¹ at a flow rate of 5 μ l min⁻¹ in SPR-buffer (10 mM HEPES, pH 7.4, 150 mM NaCl, 5 mM MgCl₂) to the preformed phospholipid layer. After 10 min protein solutions were replaced by SPR-buffer and signal decrease was observed for 30 min.

For binding of cRaf-RBD to H-Ras-7HDFar inserted into the lipid monolayer, nucleotide exchange with EDTA and GDP or the non-hydrolysable GTP analogue GppNHP- (5'-guanosyl-[β , γ -imido]-triphosphate) was performed on lane as described¹³. Net kinetics (binding of RBD to Ras*GppNHP-binding to Ras*GDP) were calculated for 250 nM of wild-type RBD and the mutant RBD^{R89L}. Flow rates were kept at 5 μ l min⁻¹ in all experiments.

Transformation assays

PC12 cell culture and microinjections were performed as described¹⁵. Ras proteins were diluted to 150 μ M in PBS containing 10 μ M fluorescein dextran marker. Forty hours after injection the transformed cells showed a typical differentiated phenotype with neurite-like outgrowth. Cells with neurites longer than three times the cell diameter were scored positive.

Received 13 August; accepted 21 October 1999.

- Resh, M. D. Regulation of cellular signalling by fatty acid acylation and prenylation of signal transduction proteins. *Cell. Signal.* **8**, 403–412 (1996).
- Wittinghofer, A. Signal transduction via Ras. *Biol. Chem.* **379**, 933–973 (1998).
- Gelb, M. H. Protein prenylation, et cetera: signal transduction in two dimensions. *Science* **275**, 1750–1751 (1997).
- Waddick, K. G. & Uckun, F. M. Innovative treatment programs against cancer. I. Ras oncoprotein as a molecular target. *Biochem. Pharmacol.* **56**, 1411–1426 (1998).
- Page, M. J. et al. Expression and characterization of the Ha-ras p21 protein produced at high levels in the insect/baculovirus system. *J. Biol. Chem.* **264**, 19147–19154 (1989).
- Schmittberger, T., Cotté, A. & Waldmann, H. Synthesis of characteristic lipopeptides of lipid modified proteins employing the allylester as protecting group. *Chem. Comm.*, 937–938 (1998).
- Nagele, E., Schelhaas, M., Kuder, N. & Waldmann, H. Chemoenzymatic synthesis of N-Ras lipopeptides. *J. Am. Chem. Soc.* **120**, 6889–6902 (1998).
- Tucker, J. et al. Expression of p21 proteins in *Escherichia coli* and stereochemistry of the nucleotide-binding site. *EMBO J.* **5**, 1351–1358 (1986).
- Bordier, C. Phase separation of integral membrane proteins in Triton X-114 solution. *J. Biol. Chem.* **256**, 1604–1607 (1981).
- Kalb, E., Frey, S. & Tamm, L. K. Formation of supported planar bilayers by fusion of vesicles to supported phospholipid monolayers. *Biochim. Biophys. Acta* **1103**, 307–316 (1992).
- Schelhaas, M. et al. Chemoenzymatic synthesis of biotinylated ras peptides and their use in membrane binding studies of lipidated model proteins by surface plasmon resonance. *Chem. Eur. J.* **5**, 1239–1252 (1999).
- Herrmann, C., Martin, G. A. & Wittinghofer, A. Quantitative analysis of the complex between p21ras and the Ras-binding domain of the human Raf-1 protein kinase. *J. Biol. Chem.* **270**, 2901–2905 (1995).
- Lenzen, C., Cool, R. H., Prinz, H., Kuhlmann, J. & Wittinghofer, A. Kinetic analysis by fluorescence of the interaction between Ras and the catalytic domain of the guanine nucleotide exchange factor Cdc25Mm. *Biochemistry* **37**, 7420–7430 (1998).
- Bar-Sagi, D. & Feramisco, J. R. Microinjection of the ras oncogene protein into PC12 cells induces morphological differentiation. *Cell* **42**, 841–848 (1985).
- Schmidt, G. et al. Biochemical and biological consequences of changing the specificity of p21ras from guanosine to xanthosine nucleotides. *Oncogene* **12**, 87–96 (1996).
- Willumsen, B. M., Cox, A. D., Solski, P. A., Der, C. J. & Buss, J. E. Novel determinants of H-Ras plasma membrane localization and transformation. *Oncogene* **13**, 1901–1909 (1996).
- Siddiqui, A. A., Garland, J. R., Dalton, M. B. & Sinensky, M. Evidence for a high affinity, saturable, prenylation-dependent p21Ha-ras binding site in plasma membranes. *J. Biol. Chem.* **273**, 3712–3717 (1998).
- Shahinian, S. & Silvius, J. R. Doubly-lipid-modified protein sequence motifs exhibit long-lived anchorage to lipid bilayer membranes. *Biochemistry* **34**, 3813–3822 (1995).
- MacDonald, R. C. et al. Small-volume extrusion apparatus for preparation of large, unilamellar vesicles. *Biochim. Biophys. Acta* **1061**, 297–303 (1991).
- Mayer, L. D., Hope, M. J. & Cullis, P. R. Vesicles of variable sizes produced by a rapid extrusion procedure. *Biochim. Biophys. Acta* **858**, 161–168 (1986).
- Schelhaas, M., Glomsda, S., Hänsler, M., Jakubke, H. D. & Waldmann, H. Enzymatic synthesis of peptides and Ras lipopeptides employing choline ester as a solubilizing, protecting, and activating group. *Angew. Chem. Int. Edn Engl.* **35**, 106–109 (1996).

Acknowledgements

We thank C. Nowak for technical support, D. Vogt for cloning C-terminally truncated H-Ras and H. Prinz for mass spectroscopic analysis. This research was supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

Correspondence and requests for materials should be addressed to H.W. (e-mail: herbert.waldmann@mpi-dortmund.mpg.de) for chemoenzymatic synthesis of lipopeptides and lipoproteins and to J.K. (e-mail: Juergen.kuhlmann@mpi-dortmund.mpg.de) for protein synthesis and coupling, biophysical and biological experiments.

Quick-fire genes, pipettes, RNA

"It was a dark and stormy night ..." Nalgene have just the thing for that first line.

Biochemicals range

Cayman Chemical www.caymanchem.com

Latest edition of this biochemicals catalogue

Hundreds of new biochemicals are listed in Cayman's 1999–2000 catalogue, including eicosanoids, nitric-acid reagents, lipids, fatty acids, inhibitors, antibodies and enzymes. Also included are assay kits for eicosanoids, cyclic nucleotides, cytokines, hormones and nitrate/nitrite. Cayman also promise many of their "popular underwater photos".

Reader Service No. 100

MicroSpot 2500

BioRobotics www.biorobotics.co.uk

Convenient and quick microarray dispensing

These new microarray spotting pins are designed for rapid microarray printing, at high feature density and with less waste of valuable DNA. The stainless-steel pins have a quill design with a fluid reservoir, making it possible to dispense more than 80 spots onto coated glass slides from a single visit to the source microplate. MicroSpot 2500 pins, which have a tip diameter of 100 μm , can produce microarray features of 140 μm diameter.

Reader Service No. 101

Weatherproof lab notebook

Nalge Nunc www.nalgenunc.com

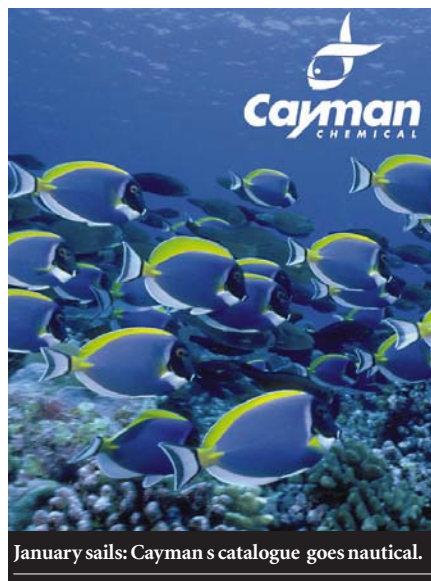
Rain or shine, keep taking those notes

Nalgene's new Field Notebook is designed for record-keeping in harsh or wet environments where paper notebooks would be impractical. Each Field Notebook comprises 96 moisture-resistant Polypaper plastic pages with quarter-inch grid lines. Durable plastic covers provide a permanent binding that makes records virtually indestructible, and all Nalgene Field Notebooks are spiral-bound, so pages lie flat to make record-keeping convenient.

Reader Service No. 102



Go buy the book: Nalgene's Field Notebook.



January sails: Cayman's catalogue goes nautical.

Template Generation System

Finnzymes Oy www.finnzymes.fi

Transposon-mediated insertion of primer binding sites for sequencing

The Template Generation System (TGS) provides the tools for inserting the artificial Mu transposon into foreign DNA at random locations in a simple *in vitro* reaction. TGS can be used to provide primer binding sites for: shotgun sequencing of large clones directly without subcloning, rapid sequencing of ordered templates without primer walking, and random insertional mutagenesis of cloned DNA.

Reader Service No. 103

arraySCOUT

LION Bioscience www.lion-ag.de

Software tool for analysis of DNA-array data

arraySCOUT was developed for the rapid identification of gene clusters and for interpreting subtle expression differences in multiple expression-profiling experiments. It uses SRS, LION's proprietary database query system, and can be used to interpret gene-expression data independently of the hardware platform that is used. The SRS link provides direct access to functional annotations of genes with interesting expression patterns.

Reader Service No. 104

Spreadex gel kits

Elchrom Scientific www.elchrom.com

Gels for demanding separations such as microsatellite analysis

The Spreadex gel matrix for DNA gel electrophoresis has a claimed resolving power

three times higher than conventional polyacrylamide gels. Seven kits are available to allow users to prepare and run 20 high-resolution mini gels. Each is optimized to resolve DNA fragments in a different size range. The gels cover seven DNA size ranges from 50 to 800 bp, and can achieve single-use pair resolution up to 200 bp on 5-cm-long gels.

Reader Service No. 105

Steam sterilization

Hotpack www.hotpack.com

Steamy publication from Hotpack

This new steam sterilization system, featured in a new colour brochure, is designed for precise liquid sterilization via the use of a load probe and F_0 programming. With this programming feature the risks of boil-over or media degradation are kept to a minimum. The Hotpack Laboratory Steam Sterilizer is controllable to $\pm 0.5^\circ\text{C}$.

Reader Service No. 106

Mini-UniPrep

Whatman www.whatman.com

Syringeless filters for rapid sample preparation

The Mini-UniPrep syringeless filter is claimed to remove particulates from samples in a third of the time taken by syringe-coupled devices and autosampler vials. After sample preparation, the Mini-UniPrep can be placed in most autosamplers designed for 12 $\times$ 32 mm vials. UniPrep consists of two parts: a chamber and a plunger. A filtration membrane is located at one end of the chamber and a pre-attached cap/septum at the other. By pressing the plunger through the liquid, the filtrate is forced into the plunger reservoir. Air escapes via the vent hole until the locking ring is engaged, forming a liquid-tight seal.

Reader Service No. 107



Take the plunge: Mini-UniPrep needs no syringe.

TRI Reagent

Molecular Research Center www.mrcgene.com

An improved version of an established single-step method for total RNA isolation

This RNA isolation technique offers simultaneous isolation of RNA, DNA and protein from a single biological sample (*Biotechniques* 15, 532–537; 1993). This allows for the comprehensive analysis of gene expression in a variety of samples of human, animal, plant and viral origin. RNA isolation is complete in less than an hour, and DNA and protein isolations in less than 3 hours.

Reader Service No. 108

3D Flat Scanner

Nanonics Imaging Ltd www.nanonics.co.il

Scanning for optical microscopes

The Nanonics 3D Flat Scanner is an ultra-thin (7 mm) three-dimensional scanner, that can be transparently placed on all optical microscope mechanical stages. The scanner features a large, clear, central aperture making it suitable for confocal and CCD-based optical sectioning, 4 Pi and nonlinear far-field optical imaging, near-field optical imaging, AFM imaging, and optical tweezer applications. Fine, sub-ångström resolution over 100 µm in x, y and z directions is possible with additional electronic rough scanning of many millimetres in x and y with 1 µm resolution in order to find the region of interest. Sensors are available for scan correction and high-resolution monitoring of the scanner position with closed-loop operation.

Reader Service No. 109



Stage craft: Nanonics' transparent scanner.

Fast pipettor

Bel-Art Products

www.bel-art.com

A single-handed approach to fast pipetting

The key to the single-handed operation of the Scienceware Fast Release Pipette Pump II Pipettor is a collar that gives a secure hold on both glass and plastic pipettes. The pump's knurled thumb wheel allows the user to aspirate or dispense the exact volume required. When pressure is applied to the "fast release lever", the liquid immediately drains. Three colour-coded versions are available: 2 ml is blue, 10 ml green, and 25 ml red.

Reader Service No. 110

TransformAid

Fermentas

www.fermentas.com

A novel approach to bacterial transformation

The TransformAid Bacterial Transformation Kit eliminates the need for prepared competent cells. With this kit it is possible to go from overnight bacterial culture to transformed cells in less than an hour, or from bacterial culture colonies on LB plates to transformed cells in 2.5 hours. High transformation efficiency (over 10^7 colonies per mg plasmid DNA) and excellent reproducibility are claimed for suitably prepared cultures of most bacterial strains commonly used for routine cloning.

Reader Service No. 111

GeneStorm

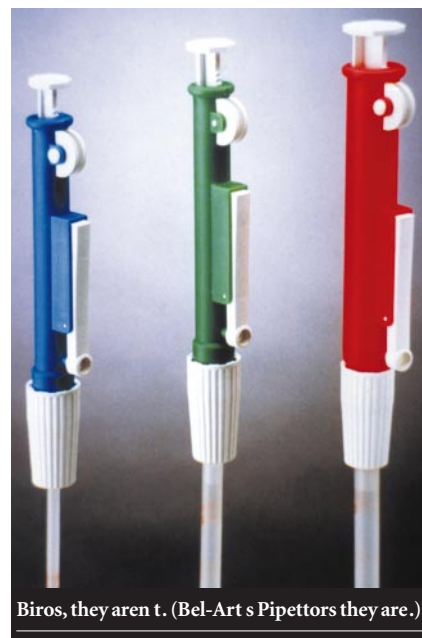
Invitrogen

www.invitrogen.com

Expression-ready mouse genes

These expression-ready clones are full-length mouse open reading frames cloned into the mammalian expression vector, pcDNA3.1/GS. Each clone is expression-tested, and completely sequenced. The pcDNA3.1/GS vector contains tags to simplify detection and purification of the expressed mouse genes. The GeneStorm clone collection is organized into categories that include signal transduction, cytokines, and cell receptors.

Reader Service No. 112



Biros, they aren't. (Bel-Art's Pipettors they are.)

SterilGARD 2.5

Baker

www.bakerco.com

A compact bench-top biological safety cabinet

The SterilGARD 2.5 vertical-flow safety cabinet controls airborne particulate contaminants and provides protection to personnel, the product and the environment. The cabinet, for sterile product preparation and biological experimentation involving agents of low and moderate risk, incorporates Baker's high-velocity return air slot design, and comes equipped with HEPA supply filter. When exhausted to the room, the SterilGARD 2.5 is classified as a Class II, Type A cabinet; properly vented to the outdoors through a facility exhaust system, the SterilGARD 2.5 exceeds minimum standards for a Class II, Type B3 cabinet.

Reader Service No. 113

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

RT activity

the best assays
www.cavidi.com



READER ENQUIRY NO. 2

Check out our homepage at <http://www.resgen.com>

BACs
YACs
PACs
cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 2

Biomolecular Synthesis oligos, synthetic peptides and polyclonal antibodies

Contact Severn Biotech Ltd
Tel: 01562 825 286
Fax: 01562 825 284
email: info@severnbiotech.com
www.severnbiotech.com

READER ENQUIRY NO. 2